TWIST1 homodimers and heterodimers orchestrate lineage-specific differentiation

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

The extensive array of bHLH transcription factors and their combinations as dimers underpin the diversity of molecular function required for cell type specification during embryogenesis. The bHLH factor TWIST1 plays pleiotropic roles during development. However, which combinations of TWIST1 dimers are involved and what impact each dimer imposes on the gene regulation network controlled by TWIST1 remain elusive. In this work, proteomic profiling of human-TWIST1 expressing cell lines and transcriptome analysis of mouse cranial mesenchyme have revealed that TWIST1 homodimer and heterodimers with TCF3, TCF4 and TCF12 E-proteins are the predominant dimer combinations. Dimers formation or their balance are altered by disease-causing mutations in TWIST1 helix domains, which may account for the defective differentiation of the craniofacial mesenchyme observed in patients. Functional analyses of the loss and gain of TWIST1-E-protein dimer activity have revealed previously unappreciated roles in guiding lineage differentiation of embryonic stem cells: TWIST1-E-protein heterodimers activate the differentiation of mesoderm and neural crest cells which is accompanied by epithelial-to-mesenchymal transition, while TWIST1 homodimers maintain the stem cells in a progenitor state and block entry to the endoderm lineage.

Short title: TWIST1 dimers in lineage differentiation

Keywords: bHLH factor/ TWIST1/ E-protein/ embryonic stem cells/ lineage differentiation

Introduction

The helix-loop-helix (HLH) superfamily is an ancient group of transcription factors that appeared in unicellular organisms >600 million years ago and has continuously evolved and diversified since then to carry out ever-elaborating activities required for lineage specification (Murre, 2019). Through dimerization with different protein partners, bHLH factors drive the specification of many cell types during neurogenesis, haematopoiesis and myogenesis (Schlaeger *et al.*, 2004; Belle and Zhuang, 2014; Comai and Tajbakhsh, 2014; Dennis *et al.*, 2019). Eukaryotic 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 are generally present in specific cell lineages, for example MYOD in myogenic cells, NEUROD in neurogenic cells and TWIST1 in mesenchymal tissues (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 the ID proteins, that compete for E-protein binding (Benezra *et al.*, 1990). It was hypothesised that the bHLH dimer composition is regulated by the expression level of each individual bHLH factor, their level of phosphorylation and their proportion relative to each other (Centonze *et al.*, 2004; Firulli *et al.*, 2005; Firulli *et al.*, 2007). However, the diversity of dimer combinations and their functional specificity during development remains enigmatic.

46 The bHLH factor TWIST1 is highly expressed in cranial mesoderm- and neural crest- derived
47 mesenchyme (Brault *et al.*, 2001; Bildsoe *et al.*, 2009; Bildsoe *et al.*, 2013) where it is critical
48 for craniofacial development. *Twist1*$^{+/-}$ mice display craniosynostosis (Carver *et al.*,
49 2002; Connerney *et al.*, 2006) that partly phenocopies skeletal defects associated with
50 *TWIST1* haploinsufficiency in human Saethre-Chotzen Syndrome (SCS, AHC [MIM: 101400]).
51 Conditional ablation of *Twist1* in the cranial mesoderm (CM) or in the cranial neural crest
52 (CNC) leads to malformations of the cranium, facial skeleton, brain, cranial nerves and
53 muscles (Chen and Behringer, 1995; Soo *et al.*, 2002; Ota *et al.*, 2004). At the cellular level,
54 *Twist1* is required for maintaining the mesenchymal cell morphology and their potency for
55 osteo-, chondro-, and adipogenesis (Brault *et al.*, 2001; Bildsoe *et al.*, 2009; Miraoui *et al.*,
56 2010; Bildsoe *et al.*, 2013). Previous studies have highlighted the differential functions of
57 TWIST1 dimers in the osteogenic differentiation of the cranial sutural mesenchyme
58 (Connerney *et al.*, 2006; Connerney *et al.*, 2008), which is mediated by their targeted action
59 on FGF signalling (Guenou *et al.*, 2005; Rice, 2005; Miraoui *et al.*, 2010). For example, the
60 TWIST1-TCF3 heterodimer promotes mesenchymal stem cell proliferation, while
61 TWIST1-homodimer activates *FGFR2*, *OCN* and *BSP* expression for ossification. Identifying
62 TWIST1 dimerization partners and their transcriptional targets in mesenchyme will therefore
63 allow for a better understanding of the mechanisms of development regulated by TWIST1
64 and bHLH factors dimers.

65 In this study, the diversity and expression of dimerization partners of TWIST1 were
66 determined by mass-spectrometry analysis, following immunoprecipitation of
67 human-TWIST1 (hTWIST1) from mesenchymal cells, and cross-compared with *Twist1*
68 co-expression analysis in mouse embryonic head tissues. The bimolecular fluorescence

complementation (BiFC) assay was used to elucidate the balance between hetero- and homo-dimerization and to assess the potential impact of pathological mutations. Finally, to dissect the specific functions of each TWIST1-dimer and their immediate downstream targets, we genetically engineered embryonic stem cells (ESCs), in which the expression of different TWIST1-E-protein dimers could be tightly controlled, and analysed their ability to differentiate and migrate. By delineating TWIST1 molecular interactions, our work has revealed previously unappreciated layers of control in lineage determination and cellular behaviour: TWIST1-E-proteins heterodimers promote mesoderm and neural crest differentiation through epithelial-mesenchymal transition (EMT) while TWIST1 homodimer maintains a progenitor-like state and blocks entry to the endoderm lineage. Together, using recent quantitative approaches and engineered cell models, this study has generated new insights into an ancient group of bHLH factors, the regulation of their dimerization activity and their role in fine-tuning lineage specification and differentiation.

Results

Identification of bHLH partners of TWIST1 in the embryonic head mesenchyme

In order to identify potential candidates dimerizing with TWIST1 protein, we first focussed on genes co-expressed with *Twist1* in vivo by investigating tissues of the mouse embryonic head. Microarray analysis of CNC and CM cells sorted from heads of embryonic day (E)9.5 embryos of *Wnt1-Cre:GFP* and *Mesp1-Cre:GFP* transgenic mice, respectively (Bildsoe *et al.*, 2016; Fan *et al.*, 2016), revealed that 58 out of 158 known bHLH factors (Skinner *et al.*, 2010) were expressed in the head mesenchyme (Supplemental Table S1). Twelve bHLH factors were significantly enriched in CNC or CM (Fig. 1A) and 46 were expressed in both tissues (Fig.

91 1A and Supplemental Table S1). Based on their known roles in craniofacial development,
92 seven candidates were selected for validation including SIM2, TCF4, EBF1, EBF3, TAL1,
93 TWIST2 and TCF3 (isoform of E2A, a known TWIST1 partner as the positive control).
94 HA-tagged proteins (including HA-tagged GFP as negative control) expression constructs
95 were transfected into Madin-Darby Canine Kidney (MDCK) cells that stably over-express
96 hTWIST1 (referred thereafter as MDCK/hTWIST1-OE cells) and have previously been used to
97 investigate the role of TWIST1 to induce mesenchymal phenotypes (Xue and Hemmings,
98 2013; Bildsoe *et al.*, 2016). These factors were co-immunoprecipitated with TWIST1.
99 Reciprocally, TWIST1 was co-immunoprecipitated with EBF1, EBF3, TCF4 and, to a lesser
100 extent, TWIST2, SIM1 and TAL1 (Fig. 1B), demonstrating that all these factors can complex
101 with TWIST1 and are potential dimer partners. Furthermore, analysis by *in situ* hybridization
102 revealed that genes encoding EBF1, EBF3 and TCF4 were co-expressed with *Twist1* in the
103 frontonasal tissues, the cranial mesoderm and the first branchial arch of the mouse embryo
104 (Fig. 1C).

105 <u>Proteomic screening of MDCK/TWIST1-OE cells revealed TCF factors as the prevalent</u>
106 <u>partners of TWIST1</u>

107 Complementary to the above approach, TWIST1 dimer partners were identified through
108 unbiased proteomic screening using affinity purification of the TWIST1-complexes from
109 MDCK/hTWIST1-OE cells followed by mass spectrometry (Fig. 1D). Three biological replicates
110 of immunoprecipitation using an α-TWIST1 antibody were performed in parallel with mouse
111 IgG control. 846 proteins were identified and divided into three sets using the
112 ranked-product method for differential analysis ((Breitling *et al.*, 2004), Fig. 1E and
113 supplemental Table S2). "Set 1" represents 103 identified proteins that specifically bound to

114 TWIST1, "Set 2" are the proteins that only bound to IgG (377) and "Set 3" contains proteins
115 bound to both IgG and TWIST1 (362). As expected, TWIST1 was one of the top enriched
116 proteins in "Set 1" (Fig. 1E). Only two other bHLH factors, TCF12 and TCF4, were also highly
117 enriched. The rest of "Set 1" TWIST1 putative partners included RNA-binding proteins and
118 transcription factors such as Cyclin T2, TRIP6, GATA1, FHL2 and CRTC2 (Supplemental Table
119 S2). *Tcf4 and Tcf12*, like *Tcf3,* are expressed in the facial primordia and within *Twist1*
120 expression domains in the frontonasal, the branchial arches and the mesoderm in the
121 embryonic head (Fig. 1C). To recapitulate early development *in vivo*, we engineered mouse
122 ESCs in which TWIST1 expression is inducible and differentiated these ESCs into embryoid
123 bodies (EBs). Co-immunoprecipitation experiments performed with these EBs demonstrated
124 that TWIST1 interacted with endogenous TCF4 and TCF12 (Fig. 1F).

125 <u>Mutations of conserved residues in bHLH domain impaired TWIST1-TCF12 dimer formation</u>
126 <u>and function</u>

127 The TWIST1 bHLH dimerisation domain shows 84–100% sequence identity among vertebrate
128 orthologues (human, mouse, frog and zebrafish; Fig. 2A). It is also highly conserved (~70%
129 similarity) among several TWIST1 family proteins: TWIST1, TWIST2, HAND1 and HAND2,
130 suggesting functional importance (Castanon and Baylies, 2002; Qin *et al.*, 2012). There is a
131 frequent incidence of clinically significant point mutations (63 reported mostly in SCS
132 patients) in the bHLH domain of *TWIST1* ((Fig. 2A; Human genomic mutation database;
133 (Stenson *et al.*, 2003; Centonze *et al.*, 2004; Firulli *et al.*, 2005; Firulli *et al.*, 2007; Barnes and
134 Firulli, 2009). Some of these mutations are located at or around the threonine and serine
135 phosphosites (Fig. 2A) and could inhibit TWIST1 phosphorylation (Firulli *et al.*, 2005; Firulli
136 and Conway, 2008). It was previously shown that TWIST1 SCS mutations in the basic-helix I

137 region, such as R118H, may impair phosphorylation at T121 and S123 residues and alter the
138 preference of dimerization with TCF3 and HAND2 (Firulli *et al.*, 2005).

139 To validate the putative interactions identified above and test the hypothesis that TWIST1
140 pathological mutations in the bHLH region may shift the balance of homo- versus
141 hetero-dimerizations, we employed the BiFC assay, a bimolecular complementation system
142 which allows the detection of direct physical interaction between two partners by
143 fluorescence (Hu and Kerppola, 2003). A subset of mutations with known impact on the
144 basic-helix I region was selected (Fig. 2A, red): truncation mutation in the bHLH domain
145 (S123X), phosphorylation-inhibitory mutations (S123A and R118H) and phospho-mimetic
146 mutations (T121E and S123E). We also selected previously uncharacterised mutations in
147 helix II (T148A, A152P and I156N) that affect conserved residues among TWIST-family
148 members. When examined by 3D *in silico* modelling of the bHLH regions of a TWIST1/TCF12
149 dimer in the presence of DNA, the phospho-regulatory residues appear to be in close
150 proximity to the DNA interface rather than the site in contact with the E-protein (Fig. 2Bi)
151 while clinical variants on helix II affect residues located at the protein interaction interface
152 (Fig. 2Bii).

153 We tested the impact of these *TWIST1* mutations on TWIST1 ability to dimerize with TCF12,
154 which is also associated with SCS and other craniofacial anomalies (Sharma *et al.*, 2013).
155 TWIST1 and TCF12 were conjugated with different split CFP or GFP fluorescent domains (N-
156 or C-terminal: CrN173 and VN173, and CC155, respectively) and co-expressed in C3H10T1/2
157 cells where dimerization was monitored by fluorescence (Fig. 2C). The interaction of
158 TWIST1-CC155 with TCF12-CrN173 versus TWIST1-VN173 was measured by the ratio of
159 fluorescent intensity from the two different fluorophores (Fig. 2D, E). When expressed at

160 comparable levels (Fig. 2F), TWIST1 mutants displayed significant changes in their
161 preference to form homo- (GFP) or heterodimers (CFP; Fig. 2D, E). Among all mutations
162 tested, truncation at the bHLH domain (S123X) severely disrupted dimerization, while loss of
163 the N-terminus did not affect dimerization (Fig. 2E and supplemental Fig. S1A).
164 Phosphorylation-inhibitory mutations S123A and R118H increased the preference for
165 heterodimer formation, whilst the phospho-mimetic mutant (TSE for T121E & S123E) of
166 TWIST1 favoured homo-dimerization (Fig. 2E). Finally, all three missense mutations at helix II
167 compromised more severely heterodimerization (Fig. 2E).

168 Mutations associated with significant shift in dimer balance were further tested for their
169 potential impact on cell migration, a known TWIST1-driven biological process. Cell migration
170 was tracked and quantified after a scratch was made on a confluent monolayer of
171 mesenchymal stem cells transiently expressing TCF12 and different variants of TWIST1 (Fig.
172 2G). Mutations in helix I showed no effect on the cells, but cells expressing TWIST1 with
173 helix II mutations (A152P, I156N) showed reduced cell migration (Fig. 2G-H). Altogether,
174 these results demonstrate that residues in helix I and helix II are important determinants for
175 dimerization, and that disrupting TWIST1/TCF12 dimerization with mutations at the
176 interaction site in helix II may compromise mesenchymal migration.

177 Loss of *Twist1* impairs exit from pluripotency and cell fate specification

178 During ESCs differentiation, the expression of endogenous *Twist1* was first detected from
179 Day 3 and then peaked at Day 6-7 (Fig. 3A). We first generated and studied *Twist1*
180 loss-of-function (LOF) ESCs to appreciate the overall function of the ensemble of TWIST1
181 dimers during differentiation. Mono- and bi-allelic *Twist1* knockout mouse ESCs were

182 engineered using the CRISPR-Cas9 methodology (Fig. 3B, C), and the expression of
183 pluripotency and lineage gene markers was analysed (Fig. 3D-H). Pairwise statistical tests
184 revealed that both full and partial loss of *Twist1* disrupted lineage specification in a gene
185 dosage-dependent manner. Genes associated with pluripotency such as *Dppa2, Nanog*, *Sox2*
186 and *Pou5f1* were reduced in wildtype parental cell line at Day 3, but expression was retained
187 at the pre-differentiated levels in the mutant cells (Fig. 3D). Concurrently, ectodermal and
188 mesodermal genes failed to be activated in the mutant cells (Figure 3E, F). Expression of
189 EMT genes such as *Pdgfra, Snail2* and *Vegfa* were also reduced in the mutant cells while the
190 epithelial marker, *Cdh1*, was up-regulated (Fig. 3F). In contrast, endodermal markers such as
191 *Gsc, Rbm47, Foxa2* and WNT signalling components (ligand: *Wnt3*, receptors: *Fzd10* and
192 *Prickle1,* target gene: *Axin2*) were up-regulated in the mutant cells (Fig. 3G, H). These
193 functions of *Twist1* in selectively directing lineage differentiation and maintaining the
194 mesenchymal progenitors are likely to be orchestrated by a combination of TWIST1 dimers.

195 <u>Comparison of the target specificity that differs between different TWIST1 dimers</u>

196 To deconstruct the function of each TWIST1 homodimer and heterodimers during lineage
197 differentiation, a gain of function (GOF) approach was taken to analyse the downstream
198 genes controlled by individual dimers. In *Drosophila*, tethered Twist-Twist and Twist-Dorsal
199 dimers could recapitulate the function of dimers formed naturally by Twist1 and Dorsal.
200 Twist homodimer induced *Mef2* expression and muscle differentiation *in vivo*, whereas the
201 Twist-Dorsal dimer antagonised this activity (Castanon and Baylies, 2002). Following the
202 same strategy, we tethered TWIST1 and TCF proteins into dimers via flexible poly-glycine
203 linkers: TWIST1-TWIST1 (TT), TWIST1-TCF3 (T3), TWIST1-TCF4 (T4) and TWIST1-TCF12 (T12),
204 and expressed them in ESCs (Fig. 4A-C). The expression constructs were stably integrated

205 into a designated genomic locus of the A2LoxCre ESCs, downstream of a doxycycline
206 inducible promoter (Mazzoni *et al.*, 2011; Iacovino *et al.*, 2014; Sibbritt *et al.*, 2019). ESCs
207 were differentiated for three days into EBs (Fig. 4B) and assayed for the expression of
208 developmental/lineage marker genes. Before doxycycline induction, no expression of the
209 transgene was detected (Supplemental Fig. S2A). Dimer expression was induced at Day 2 of
210 differentiation in synchrony with the onset of expression of endogenous *Twist1* (see Fig. 3A).
211 To ensure that the effects of each dimer was not confounded by dissimilar level of
212 expression, different doses of doxycycline were tested and the optimal concentration was
213 selected for each cell line to achieve comparable levels of expression between each
214 construct (Fig. 4C and supplemental Fig. S2A).

215 The transcriptomes of EBs differentiated for 3 days expressing comparable levels of dimers
216 were then analysed by RNA-Seq and compared to the transcriptome of the parental
217 A2loxcre cell line (A2LC) as well as the transcriptome of a cell line expressing TWIST1
218 Monomer (T0) that were both also differentiated for 3 days (Fig. 4D-H and supplemental Fig.
219 S3). Unsupervised principle component analysis (PCA) and hierarchical clustering confirmed
220 groupings of replicates (Fig. 4D, E). Global gene-expression profile separated T3, T4 and T12
221 from TT, T0 and A2LC EBs (Fig.4E, F), a result which was mainly driven by genes upregulated
222 in the heterodimers. Among the three heterodimers, cells expressing T4 and T12 dimer
223 exhibited more similar gene expression profile (Fig. 4E). Differentially-expressed-genes
224 (DEGs) analysis was performed for T0, TT, T3, T4 and T12 against the A2LC cells. Examination
225 of the upregulated gene groups revealed that while all dimers activated cell migration
226 program, the heterodimers specifically activated the expression of genes associated with
227 pattern specification, mesenchymal tissues development (muscle, skeletal, connective tissue

228 etc.) and TGFβ signalling (Fig. 4G). A four-way comparison between TT or heterodimers
229 against A2LC also demonstrated a stronger ability of the heterodimers to activate
230 developmental markers, especially mesenchyme-related genes (Figure 4H; Supplemental Fig.
231 S3).

232 It was previously demonstrated by *in vitro* mobility shift assay that different dimers
233 preferentially bind different DNA sequences (Firulli *et al.*, 2007; Chang *et al.*, 2015).
234 Sequence motif analysis was performed on the proximal regulatory sequences for
235 dimer-specific DEGs. For each group, 100-600 genes were up-regulated, and 100-200 were
236 down-regulated (adjusted p-value $\leq$ 0.05, fold change $\geq$ 2; Fig. 5A; Supplemental Table
237 S3). Top ranking regulatory motifs for heterodimers-activated promoters contained a single
238 canonical E-box “CATCTG” or “CAGCTG" (Fig. 5B), which is consistent with TWIST1 ChIP-seq
239 and dimers mobility shift assays results (Firulli *et al.*, 2007; Chang *et al.*, 2015). TT targets
240 were uniquely enriched in double E-boxes with a five nucleotide spacer as previously
241 reported by ChIP-Seq in human mammary epithelial cells over-expressing *TWIST1* (Chang *et*
242 *al.*, 2015).

243 To further dissect the functionality of each dimer, we performed additional enrichment
244 analyses focusing on development related gene sets and taking into account gene
245 expression changes relative to the parental A2LC cells. Neural-ectoderm differentiation
246 derivatives were repressed by the monomer while mesoderm and endoderm cell fates were
247 repressed by TT (Supplemental Fig. S4A; Supplemental Fig. S3; Supplemental Table S3). On
248 the other hand, heterodimers activated overlapping sets of genes associated with
249 mesoderm and neural ectoderm development (Supplemental Fig. S4A; Supplemental Table
250 S3). T3 dimer activated heart developmental processes, including heart vasculature, muscle,

251 and blood cells, while the most frequent processes induced by T4 were related to neuron
252 differentiation (Supplemental Fig. S4A; Supplemental Table S3). T12 activity was
253 intermediate between T3 and T4. A comparison down to key cell lineage and pathway genes
254 also reflected these changes (Fig. 5C, D) which were furthermore validated by RT-qPCR
255 experiments (Figure 5E). The most evident changes in the heterodimer groups were the
256 downregulation of pluripotency genes and a burst of expression of lineage markers,
257 especially those for mesoderm and neural crest ectoderm development (Fig. 5C, E).
258 Endoderm markers were repressed by all dimers except T3 (Fig. 5D, E; Supplemental Fig. S3;
259 Supplemental Table S3).

260 Taken together, our results have demonstrated that TWIST1 dimerization partners drive
261 different gene expression programs which may be influenced by the preference for specific
262 DNA motifs. Heterodimers appeared to drive ectoderm and mesoderm differentiation,
263 whereas TWIST1 homodimer repressed the expression of developmental genes while
264 maintaining the pluripotency state.

265 <u>TCF12-TWIST1 dimer activates EMT more robustly and antagonises TWIST1 homodimer</u>

266 <u>activity for critical signalling pathway genes</u>

267 *Twist1* and *Tcf12* have been shown to act synergistically for cranial tissue development
268 (Sharma *et al.*, 2013). Previous studies have suggested that homo- and heterodimers
269 regulate antagonistically the skull suture (Connerney *et al.*, 2006; Connerney *et al.*, 2008).
270 Therefore, we focused further our analysis on the T12 dimer and its ability to induce
271 transcription of developmental pathways, in comparison with TT (Supplemental Fig. S4B, C
272 and supplemental Table S4). T12 expressing cells significantly activated more genes

273 associated with TGFβ signalling (FDR 0.007), WNT signalling (FDR 0.017) and EMT (FDR 0.002;
274 around 50% of all EMT genes were enriched in T12 cells; Fig. 6B). T12 expressing cells were
275 also enriched for genes associated with processes of neuronal, skeletal, face and
276 cardiovascular system development (Fig. 6B and supplemental Fig. S4B). Examining
277 specifically genes encoding transcriptional regulators associated with craniofacial
278 development (Fig. 6B, Face development), we noted that a subset of critical developmental
279 regulators such as those associated with TGFβ signalling (e.g. *Tgfbr1, Pdgfrα*) and WNT
280 signalling (*Prickle1, Dkk1*) were antagonistically regulated by the homodimer and the
281 heterodimer (Fig. 6C). Other genes, represented by *Fgfr2, Jun, Chd7* and *Mmp2* were
282 activated to a greater extent by T12 than TT (Fig. 6C).

283 The differential expression of downstream genes between T12 and TT and the functional
284 consequences for craniofacial development are therefore likely to be the outcome of the
285 balancing activity of these two dimers for transcriptional regulation.

286

287 **Discussion**

288 <u>TCF factors are the major dimerization partner and transcriptional modulator of TWIST1</u>

289 Our study has identified TCF3, TCF4 and TCF12 E-proteins, which are expressed in the cranial
290 mesenchyme, as the predominant interacting partners of TWIST1. This is also supported by
291 the results of a TWIST1 proximity labelling screen showing that the TCF factors partner with
292 TWIST1 consistently in neural crest cells, mesenchymal stem cells and 3T3 cells (Fan et al,
293 unpublished results). E-proteins are also major dimerization partners of several other Class II

294 bHLH factors (Murakami *et al.*, 2004; Forrest *et al.*, 2014). For example, TCF3, TCF4 and
295 TCF12 are top interacting candidates of HAND2 and enhance its transcriptional activity at
296 the promoter of target genes (Murakami *et al.*, 2004). Consistent with these observations,
297 we found that TWIST1 induced differentiation programs more effectively when coupled with
298 E-proteins. Also similar with HAND2, coupling TWIST1 with different E-proteins resulted in
299 slightly altered DNA binding activity as previously shown by electromobility shift assay
300 (Murakami *et al.*, 2004; Firulli *et al.*, 2007). These observations suggest that E-proteins might
301 be the primary transcriptional modulators of TWIST1, and potentially other class II bHLH
302 factors, for the regulation of various tissue-specific programs during development.

303 <u>Pathological mutations induce bHLH factor dimerization imbalance</u>

304 For TWIST-family bHLH factors, dimerization can be modulated by phosphorylation of the
305 bHLH domain (Centonze *et al.*, 2004; Firulli *et al.*, 2005; Firulli *et al.*, 2007).
306 Hypo-phosphorylation mutations (T121A, S123A) near or at the helix I phosphorylation sites
307 were previously shown to favour homodimerization (Firulli *et al.*, 2005). Nevertheless, a
308 triple phospho-mutant model of TWIST1 (S123A, T148A and A184A) in oncogenic cell lines
309 showed preference for dimerization with TCF3 and HAND2 (Wang *et al.*, 2017). In line with
310 this, we found that two individual hypo-phosphorylation mutants (R118H, S123A) both
311 increased preference for TCF12 heterodimer formation, whilst phospho-mimetic forms at
312 T121 and S123 reduced heterodimerization. We noted that incorporation of the negatively
313 charged phosphate group to these residues may reduce binding to DNA, rather than directly
314 interfering with the dimerization interface. Therefore, T121/S123 dephosphorylated TWIST1
315 may favour heterodimerization through stabilising DNA-binding or adapting a helix
316 conformation that enhances E-protein binding. Driven by a potentially different mechanism,

317 TWIST1 helix II mutations at the interaction site severely compromised hetero- versus
318 homodimerization potential, which had greater functional consequence than helix I variants
319 and impacted on the migration of mesenchymal cells. Altogether, these observations
320 demonstrate that these residues constitute an important interface for dimer balance
321 adjustment and therefore play a critical role for cell differentiation and development.

322 RNA-seq analysis of differentiating ESCs revealed that the TWIST1-TCF12 heterodimer is a
323 potent driver of EMT and mesenchymal differentiation compared to TWIST1 homodimer.
324 TWIST1-TCF12 also induced signalling pathways and key transcription regulators associated
325 with head mesenchyme development. These genes were either inactive or repressed in cells
326 expressing the TWIST1 homodimer. The *Twist1* heterozygous (*Twist1*$^{+/-}$) mouse model that
327 recapitulates SCS phenotypes displays excessive FGF and TGFβ signalling at the osteogenic
328 front, as revealed by the expansion of FGFR2, pSmad 1/5/8, and ID protein signal
329 (Connerney *et al.*, 2006; Connerney *et al.*, 2008). These mis-regulated signalling activities
330 could therefore be a consequence of imbalanced TWIST1-E-protein activities.

331

332 <u>Dimerization-dependent activities of TWIST1 in progenitor cell lineage specification</u>

333 TWIST1 regulates multiple pathways during lineage specification. Loss of *Twist1* function
334 leads to retention of pluripotency, defective mesoderm and ectoderm differentiation, as well
335 as disrupted EMT. EMT is tightly associated with ESCs differentiation, especially the
336 specification of mesoderm (Nakaya and Sheng, 2008; Evseenko *et al.*, 2010; Iacovino *et al.*,
337 2014). Major EMT drivers including *Twist1, Zeb2, Snail1* and *Snail2* are induced by canonical
338 WNT signalling in mesendoderm differentiation (Ten Berge *et al.*, 2008), while constitutive
339 activation of WNT further directs precursor cells into primitive endoderm (Price *et al.*, 2013).

340 Both *Twist1* and *Snail1* are required for WNT induced mesoderm differentiation (This study
341 and (Sharma *et al.*, 2013) but only *Twist1* negatively feedbacks on WNT signalling regulation
342 and represses endoderm cell fates. Over-expression of the TWIST1 homodimer blocked WNT
343 activity and endoderm differentiation while loss of TWIST1 shifted differentiation towards
344 endoderm. On the other hand, ablation of *Snail1* promotes ectodermal instead of
345 mesendodermal fate without affecting WNT pathway genes (Sharma *et al.*, 2013). These
346 observations suggest a dichotomic control carried out by TWIST1 during development: 1) an
347 EMT-dependent role in promoting mesoderm and neural crest commitment and 2) an
348 EMT-independent role in endodermal repression potentially by negatively feedback on WNT
349 activity.

350 These activities undertaken by TWIST1 during lineage commitment are directed by
351 E-proteins. E-proteins display overlapping expression patterns and developmental functions
352 in embryonic tissues (Murakami *et al.*, 2004; Wang and Baker, 2015). They are expressed in
353 the cranial mesenchyme and proliferating neuroepithelium at mid-gestation (E9.5-E11.5)
354 (Murakami *et al.*, 2004). *Tcf3* and *Tcf12* expression are restricted to the proliferative
355 ventricular and sub-ventricular zones and are down-regulated during neurogenesis
356 (Uittenbogaard and Chiaramello, 2002; Pfurr *et al.*, 2017). *Tcf4* expression persists in the
357 adult brain in regions that overlap with regions expressing genes encoding neuronal bHLH
358 factors such as *Atoh1*, *Ascl1* and *Neurog2*, and initiates neuronal specification (Forrest *et al.*,
359 2014). Consistent with these findings, tethered TWIST1-TCF4 was shown to modulate the
360 activity of neural-ectoderm genes, whereas tethered TWIST1-TCF3 most strongly activated
361 the mesoderm genes. TWIST1-E-protein heterodimers activated the neural crest cell
362 differentiation and TGFβ signalling pathway, which in turn up-regulated ID proteins that

363 could compete for E-protein binding, providing a potential negative feedback mechanism to
364 modulate the activity of the heterodimers (Connerney *et al.*, 2006).

365 In conclusion, our study has uncovered the predominant dimerization partners of TWIST1
366 and their role in regulating lineage differentiation during early development. TWIST1
367 interacts with TCF proteins to promote the differentiation of neural crest and mesoderm
368 progenitors, and to repress endoderm lineage differentiation with the help of TWIST1
369 homodimer. The TWIST1-TCF dimers enable the mesenchyme precursors to acquire
370 myogenic, skeletogenic and neurogenic fates. In the context of embryogenesis, we showed
371 that the mutations of *TWIST1* in congenital diseases are likely to alter the balance of dimer
372 compositions, which may be responsible for the mesenchymal tissue defects observed in
373 adults.

374 ## Figure Legends

375 **Fig. 1. Unbiased high-throughput identification of proteins interacting with TWIST1 by**
376 **co-expression analysis, co-immunoprecipitation (IP) and mass spectrometry. A.** Relative
377 expression of genes encoding known bHLH factors in cranial neural cells (CNC) versus cranial
378 mesoderm (CM) of E9.5 embryonic head collated from transcriptome data previously
379 published (Bildsoe *et al.*, 2016; Fan *et al.*, 2016). **B.** Detection of HA-tagged proteins (left
380 blot: α-HA) after IP of TWIST1 (IP: α-TWIST1) and reciprocal IP of HA-tagged proteins (IP:
381 α-HA) and detection of TWIST1 (Blot: α-TWIST1) from lysates of hTWIST1-expressing MDCK
382 cells transfected with constructs expressing the HA-tagged protein partners, with expression
383 confirmed in the input (right blot: α-HA) **C.** Expression of candidate transcripts analysed by
384 whole mount *in situ* hybridization in regions of E9.5 embryonic head. Lateral views with
385 anterior to the left. cm, cranial mesoderm; fn, frontonasal tissue; ba, brachial arch. Scale bar:
386 500 µm. **D.** Experimental strategy for identifying TWIST1 interacting partners from
387 hTWIST1-expressing MDCK cells. TWIST1-containing protein complexes are isolated on
388 beads using an α-TWIST1 antibody before analysed by tandem mass spectrometry which
389 determines the ion spectrum of the different peptides (one example shown here). **E.** Groups
390 of proteins identified using the non-parametric rank product method based on cumulative
391 count of reporting a protein across independent experiments. Sets: red, TWIST1 IP; black,
392 non-specific IgG; yellow, common to TWIST1 IP and IgG control. Normalised unique peptide
393 spectrum match (PSM) of the detected proteins are shown on both axis for each condition.
394 The size of the dots corresponds to $-\log_{10}$ P-value. IGKC: Immunoglobulin Kappa Constant **F.**
395 Detection of TCF12 (Blot: α-TCF12) or TCF4 (Blot: α-TCF4) endogenous proteins after IP of
396 TWIST1 (IP: α-TWIST1) from lysates of embryoid bodies over-expressing TWIST1, compared

397 with results obtained for non-specific IgG and the input.

398 **Fig. 2. Human disease-causing mutations disrupt TWIST1-TCF12 dimer function. A.**
399 Alignment of selected mouse TWIST-family bHLH proteins and hTWIST1 sequences. The
400 degree of similarity of the different regions is indicated in dark shades. Blue and red bars
401 indicate missense pathological mutations reported in the Human Gene Mutation Database
402 (Stenson *et al.*, 2003). The red bars identify the mutations analysed in this study. The pink
403 squares are predicted phosphosites. **B.** *In silico* 3D structure prediction of the bHLH regions
404 of TWIST1-TCF12 dimer binding to DNA, based on the structure of NeuroD-TCF3 dimer as
405 determined by crystallography (O'Donoghue et al. 2015; http://aquaria.ws). Enlarged
406 representation of **i.** basic-helix I and **ii.** helix II domains of TWIST1 are shown. TWIST1
407 residues mutated in disease and their atomic interaction with TCF12 are highlighted in
408 yellow. **C.** Experimental setup for the BiFC assay for which three independent constructs
409 expressing TWIST1*-CC155, TWIST1-VN173 and TCF12-CrN173 were equimolarly
410 co-transfected in C3H10T1/2 cells. TWIST1*: wildtype or SCS variants. Interactions between
411 the different proteins and the relative proportion of TWIST1 heterodimers versus TWIST1
412 homodimers are then assessed by quantifying and comparing CFP and GFP fluorescence
413 intensities as shown in E. **D.** Representative confocal microscopy images of BiFC assays
414 performed for wildtype (WT) and A152P-mutant TWIST1. Signals for dimers:
415 TWIST1*-CC155/TWIST1-VN173 (T/T) or TWIST1-VN173/TCF12-CrN173 (T/E) were shown as
416 indicated. Red channel shows TWIST1*-CC155 detected by immunostaining with an α-HA
417 antibody. Cell nuclear was stained by DAPI. Scale bar: 10 µm. **E.** Heterodimer/homodimer
418 CFP/GFP ratio quantified for individual cells (each dot corresponding to one cell) in BiFC
419 assays. TWIST1 variant and number of cells assayed are indicated in chart below. TSE

420 corresponds to T121E; S123E double mutant. Red bars represent standard errors of the
421 mean. Data analysed by non-parametric Mann-Whitney U test (two-tailed); *P ≤ 0.05, ***
422 P ≤ 0.001. **F.** Western blot analysis of BiFC proteins expression in C3H10T1/2 cells. TWIST1
423 BiFC constructs were detected with an α-GFP antibody; TCF12-CRN173 was detected with
424 an α-TCF12 antibody. Loading control (β-tubulin) is also shown. #: non-specific band. **G.**
425 Example images from a scratch assay. C3H10T1/2 cells were co-transfected with
426 TWIST1*-CC155 (WT or A152P) and TCF12-CrN173 expression vectors. Time lapse images
427 were taken every 15 mins over an eight-hour-period, after a scratch was made on the cell
428 monolayer. Cells were tracked (pseudo-coloured in blue) and area of cell free gap was
429 quantified in ImageJ software. Dashed lines indicate original gap boundaries. **H.**
430 Quantification of migration (% of total area in 8 hrs), of C3H10T1/2 cells co-transfected with
431 TWIST1*-CC155 and TCF12-CrN173 expression vectors. Data represent the mean ± standard
432 errors of n=6 independent experiments, analysed by one-way ANOVA with Holm-Sidak's
433 post-test; **P ≤ 0.01, *** P ≤ 0.001 or non-significant (ns).

434 **Fig. 3. TWIST1 is required for mesoderm and ectoderm differentiation. A.** RT-qPCR analysis
435 of *Twist1* expression relative to mean value of *Gapdh* expressions (Y-axis) during ESC
436 differentiation (X-axis). Each time point represents the mean value of n=3 independent
437 experiments. **B.** *Twist1* alleles of two independent ESC clones (*Twist1*$^{+/-}$ and *Twist1*$^{-/-}$)
438 generated by CRISP-Cas9 editing. Genomic positions targeted by the gRNA, Sanger
439 sequencing chromatograms and deduced gDNA sequences are shown. Mutations are
440 frameshifts in the coding sequence of exon 1. *Twist1*$^{+/-}$ clone: allele #1 is unmodified; allele
441 #2: 19 bp deletion. *Twist1*$^{-/-}$ clone: allele #1: 2 bp insertion; allele #2: 2 bp insertion and 91

442 bp deletion. **C.** Western blot analysis of TWIST1 proteins in the clones characterised in B and
443 the parental ES wildtype (WT) cell line. Loading control (β-tubulin) is shown. **D-H.** RT-qPCR
444 analysis of the expression (Y-axis) of selected genes (X-axis) involved in cell lineage, EMT or
445 WNT signalling in mutants, wildtype (WT) and undifferentiated ESC, at day 3 of *in vitro*
446 differentiation. Expressions were normalized against the mean value of *Tbp* and *Gapdh*
447 expressions and the wildtype sample at Day 3 of differentiation. Each data points represent
448 the mean ± standard errors of n=3 independent experiments. Data analysed by
449 non-parametric Mann-Whitney U test (two-tailed). *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001.

450 **Fig. 4. Transcriptome profiles of embryoid bodies expressing different TWIST1-dimers and**
451 **monomer. A.** Schematic of doxycycline-inducible expression constructs: *Twist1-Tcf3* (T3),
452 *Twist1-Tcf4* (T4), *Twist1-Tcf12* (T12), *Twist1-Twist1* (TT) and *Twist1* only (T0; TWIST1
453 monomer) recombined into designated locus in A2loxCre parental ESCs (Mazzoni *et al.*,
454 2011; Iacovino *et al.*, 2014). Proteins in dimer are joined by a flexible poly-glycine linker
455 (black box). **B.** Experimental setup to obtain dimers expressing EBs: ESCs are differentiated
456 into EBs for 3 days. Dimer expression is induced with doxycycline 16 h before collection **C.**
457 Detection by western blot using an α-TWIST1 antibody of TWIST1 dimers and monomer in
458 the engineered ESCs and the parental A2loxCre cell line (A2LC) after doxycycline induction.
459 Loading control (β-tubulin) is also shown. **D.** Principal component analysis of the
460 transcriptomes of the dimers/monomer-expressing ESCs and A2LC EBs (n=3 independent
461 RNA-Seq for each cell line) showing the first two primary components. **E.** Correlation map of
462 the different RNA-seq data showing Euclidean distance between cells lines (from high
463 correlation in blue to low correlation in red). **F.** Scaled expression (Y-axis) of the 125 genes
464 with top PC loadings across all cell lines. Each line corresponds to one gene. **G.** Gene

465 Ontology analysis of upregulated genes (adjp < 0.05, Log2FC > 1) in all cell lines versus A2LC.
466 The Top 10 most significant non-redundant GO categories for each cell line were plotted.
467 Dot sizes represents "GeneRatio" (gene upregulated/gene in GO set) and adjusted P-value is
468 colour-coded. Number of identified genes in each group is show between parentheses. **H**.
469 Four-way plot visualization of DESeq2 result. Log2FC comparisons between TT and T12
470 treatments versus A2LC were performed. Blue nodes indicate genes significantly changed
471 for both cell lines. Triangular shapes indicate values exceeding the limit of for the plot. Size
472 change reflects the magnitude of Log2FC values. Vertical and horizontal dashed lines
473 indicate |Log2FC|= 1. Key markers and pathway genes are indicated: Black, pluripotency;
474 Green, endoderm; Red, mesoderm; Blue, ectoderm & EMT; Dark green, WNT signalling
475 pathway.

476 **Fig. 5. Differential specificity analysis of TWIST1-dimers and monomer. A.** Pairwise
477 differential gene expression analysis of the dimers/monomer-expressing ESCs against the
478 parental A2LoxCre ESCs performed with DESeq2 (Love *et al.*, 2014) (adjp < 0.05, fold change
479 ≥ 2). Number of individual and common up-regulated (Up) and down-regulated (Down)
480 genes are shown in Venn diagrams. **B.** Top three predicted binding motifs enriched in the
481 regulatory sequences (TSS +/- 2 kb) of the differentially regulated target genes of TWIST1
482 dimers, generated with HOMER (Heinz *et al.*, 2010). Names of the known motif and
483 percentage of targets containing these motifs are listed**, D.** Heatmap showing the relative
484 expression (Log_2-fold vs A2LC) of selected genes involved in cell lineage, EMT or cell
485 signalling pathways for the dimers/monomer-expressing EBs. Colour-coded normalised
486 Z-score is shown. **E.** RT-qPCR analysis of the expression (Y-axis) of key lineage markers
487 (X-axis) in different EBs. Expressions were normalized against the value of *Tbp* and the

488 wildtype sample at Day 3 of differentiation. Each data points represent the mean ±
489 standard errors of n=3 independent experiments. Data analysed by non-parametric
490 Mann-Whitney U test (two-tailed). *P ≤ 0.05, **P ≤ 0.01 or non-significant (ns). Black,
491 pluripotency markers; Green, endoderm markers; Red, mesoderm markers; Blue, neural
492 crest markers.

493 **Fig. 6. TWIST1-TCF12 heterodimer antagonises TWIST1 homodimer. A, B.** Barcode plots of
494 the enrichment of cell signalling and Developmental process-related genes in TWIST1-TCF12
495 EBs (right) versus TWIST1-TWIST1 EBs (left). Individual genes, represented by vertical bars,
496 are ranked from left to right based on $\log_2$ fold change (X-axis) for TWIST1-TCF12 EBs versus
497 TWIST1-TWIST1 EBs. The enrichment line shows the relative local enrichment (Y-axis) of the
498 bars in the plot. The dotted horizontal line indicates neutral enrichment. Values above the
499 dotted line show enrichment while values below the dotted line show depletion. **C.** Boxplots
500 depicting the expression (Y-axis) of selected genes important for craniofacial development
501 and cell signalling pathways that are enriched in TWIST1-TCF12 versus TWIST1-TWIST1
502 expressing EBs. In Red, TGFβ signalling related genes; In Green, WNT signalling related
503 genes. Dots represents replicate data points. The top and bottom edge of each box shows
504 first and third quantiles (Q1 and Q3) of the data values; the middle bar shows median values.
505 Upper whisker = highest observed value within [Q3, 1.5*Q3]; Lower whisker = lowest
506 observed value within [0.5*Q1, Q1].

507

508 ## Materials and Methods

509 **Cell lines and cell culture conditions**

510 Cell lines and their treatments are summarized in Table S6. The Madin-Darby Canine Kidney
511 (MDCK) cell line over expressing human *TWIST1* was received from Xue and colleagues (Xue
512 *et al.*, 2012). Cells were maintained in high glucose Dulbecco's modified Eagle's medium
513 (DMEM, Gibco), 10% Fetal calf serum (FCS) and 10 mM β-mercaptoethanol at 37°C and 5%
514 CO2. *TWIST1* expression was under constant selection using 4 µg/mL of puromycin (Xue *et*
515 *al.*, 2012). C3H10T1/2 cell line was received from ATCC. Cells were maintained in high
516 glucose DMEM with 10% FCS (Life Tech, Australia) and 10 mM β-mercaptoethanol at 37°C in
517 5% CO2.

518 Mouse ESCs (A2loxCre) was a gift from Kyba Lab (Lillehei Heart Institute, Minnesota, USA),
519 originally described in Mazzoni et al., (2011). The ESCs were cultured in complete mESC
520 medium (high glucose DMEM (Gibco), 12.5 % (v/v) heat-inactivated FBS (Fisher Biotec
521 Australia), 10 mM β-mercaptoethanol, 0.09 % (v/v), 1X non-essential amino acids (100X
522 Thermo Fisher Scientific), 1 % (v/v) nucleosides (100X, Merck) and 10 mil U/mL ESGRO®
523 mouse leukaemia inhibitory factor (Merck)).

524 All cell lines were tested free of mycoplasma.

525

526 **Generation of *Twist1* mutant ESCs by CRISPR-Cas9**

527 ESC editing and selection by GFP was performed as described in detail in (Sibbritt *et al.*,
528 2019). Briefly, 3µg of pSpCas9(BB)-2A-GFP (addgene plasmid #48138, a gift from Feng Zhang)
529 expressing the gRNA of interest were electroporated into 1x10^6 A2loxCre ESCs using the

530 Neon® Transfection System (Thermo Fisher Scientific). Colonies expressing GFP were picked
531 4 days after electroporation.

532 For genotyping, region surrounding the mutation site (+/- 200 bp) was amplified from cell
533 lysate, purified and cloned for sequencing. Clones with mono- or bi-allelic frameshift
534 mutations were expanded and used for subsequent experiments.

535

536 **Generation of dimer inducible ESC line**

537 The inducible TWIST1-dimer ESCs lines were generated using the inducible cassette
538 exchange method described previously (Mazzoni *et al.*, 2011; Iacovino *et al.*, 2014). The
539 protein coding sequences were cloned from mouse embryos cDNAs into the p2lox plasmid
540 downstream of the Flag-tag coding sequence (Mazzoni *et al.*, 2011; Iacovino *et al.*, 2014).
541 Plasmids were transfected into A2loxCre treated with 1 µg/mL doxycycline for 24 hrs.
542 Selection was performed in 300 µg/mL of G418 (Gibco) antibiotic for one week. Colonies
543 were then picked and tested for constructs expression following doxycycline treatment.

544

545 **Embryoid body formation**

546 Embryoid bodies (EB) were maintained in EB-medium consisting of high glucose DMEM
547 supplemented with 15% heat-inactivated FCS, 1% Non-essential amino-acid supplement and
548 10 mM β-mercaptoethanol. Single ES cells were seeded onto aggrewells in a 24 well plate
549 (Stem Cell Technologies, Cat. No. 34411) at 1.5 × 10^6 cells per well in EB-medium. Cell
550 aggregates were transferred to non-tissue culture treated 10 cm plates and EB-medium 24 h
551 later and placed onto an orbital-shaker under slight rotation at 37 °C, 5% CO2. Drug
552 treatment conditions was optimised as shown in Supplemental Fig. S2. Doxycycline was

553 supplied at 50 ng/mL for T3 and T12, 100 ng/mL for TT, and 1000 ng/mL for other cell lines
554 after 2 days in differentiation media. Cells were collected 16 hours later as expression of
555 known *TWIST1* targets peaked 12–16 hrs after induction (Supplemental Fig. S2B).

556

557 ***In situ* hybridisation**

558 The use of C57BL/6-Arc E 9.5 mouse embryos was approved by the Children's Medical
559 Research Institute/Children's Hospital at Westmead Animal Ethics Committee. Embryos
560 collection and in situ hybridisation were performed as previously described (Bildsoe *et al.*,
561 2016).

562

563 **Scratch Assay**

564 TWIST1*-CC155 (1 μg) and TCF12-CrN173 (1 μg) were co-transfected into C3H10T1/2 cells
565 seeded in wells of 6-well plate. 24 h after transfection and when cells had reached
566 confluency, a p200 pipet tip was used to make a controlled scratch on the cell monolayer.
567 Cells were then washed once with cell culture medium to remove floating cells detached by
568 the scratch. Plates were then installed into the chamber (37°C, 5% CO2) of a Cell Observer
569 Widefield microscope (ZEISS international) and bright-field images of the region scratched
570 were taken every 15 min over a 8 h period. ImageJ was used to process the imaging data.
571 Briefly, images were enhanced by subtracting the background and binary images were
572 created to assist quantification of the gap closure. Total migration from the start of imaging
573 until the time point when the first cell line closed the gap was then calculated. One-way
574 ANOVA followed by wo-tailed t-test was used to determine the significance of the
575 difference observed between the different conditions.

576

577 **Protein Affinity Purification and Mass spectrometry**

578 2x10^7 MDCK/*hTWIST1* over-expressing cells were collected for each experiment. Cell
579 pellets were thawed in 500 µl hypotonic lysis buffer (HEPES 20 mM, MgCl2 1 mM, Glycerol
580 10%, Triton 0.5%, DTT 1 mM, 1X Complete protease inhibitor [Roche], 1000 U/ml Benzo
581 nuclease) and incubated at room temperature for 15 min. Equal volume of hypertonic lysis
582 buffer (HEPES 20 mM, NaCl2 500 mM, MgCl2 1 mM, Glycerol 10%, DTT 1 mM, 1X Complete
583 protease inhibitor) was applied to the lysate. Cells were further broken down by passage
584 through gauge-25 needle 10 times and then rotated at 4°C for 30 min. After centrifugation
585 at 14,000 x g, 15 min, 450 µL lysates were incubated with 5 µg of an α-TWIST1 antibody or
586 mouse IgG at 4°C overnight with rotation. Meanwhile, Dynabead® protein G slurry
587 (Invitrogen™, cat. 10003D) was prepared by washing and blocking for 1 h at room
588 temperature in 1% BSA 0.1% triton dissolved in PBS 1X before use. 50 µL of Dynabead was
589 then added to each sample and rotated at room temperature for 30 min. Beads were then
590 washed with 500 µL ice-cold wash buffer (1:1 mixture of the lysis buffer, without nuclease) 6
591 times and transferred to new 1.5 ml tube before the last wash. Beads were then washed
592 quickly with 1 mL cold Tris·HCl pH 7.4 50 mM and then 500 µl TEAB (75 mM). Lastly, beads
593 were collected by centrifugation for 5 min at 2,000 × *g,* before being processed for
594 Mass-spectrometry analysis.

595 Tryptic digestion of the immunoprecipitated proteins was performed by adding trypsin 1:20
596 (µg) directly to the washed beads in 50 mM TEAB buffer, vortex and incubation at 37 °C
597 overnight. A second digestion was performed by adding trypsin 1:40 the next day and

598 incubation for 4 h. The tubes were then placed on a magnet and the supernatant was
599 collected and directly transferred into TFA to obtain a final concentration of 0.5%.

600 Proteolytic peptides were desalted using Oligo R3 reversed phase resin (Thermo Fisher
601 Scientific) in custom-made stage tips (Rappsilber *et al.*, 2007). Mass spectrometry (MS) was
602 performed using an LTQ Velos-Orbitrap MS (Thermo Fisher Scientific) coupled to an
603 UltiMate RSLCnano-LC system (Thermo Fisher Scientific). Raw MS data files were processed
604 using Proteome Discoverer v.1.3 (Thermo Fisher Scientific). Processed files were searched
605 against the UniProt mammalian database (downloaded Nov 2016) using the Mascot search
606 engine version 2.3.0. Searches were done with tryptic specificity allowing up to two missed
607 cleavages and a tolerance on mass measurement of 10 ppm in MS mode and 0.3 Da for
608 MS/MS ions. Using a reversed decoy database, a false discovery rate (FDR) threshold of 1%
609 was used. The lists of protein groups were filtered for first hits.

610

611 **Rank Product analysis for MS data**

612 We generated three sets of data using the non-parametric rank product method (Breitling *et*
613 *al.*, 2004) and based on how many times a protein was reported across independent
614 experiments (Supplemental Table S2). To test for cases where proteins were preferentially
615 detected (here "preferentially" is defined as present in at least 2 samples for Twist, i.e. Set
616 1, and 2 samples for Control, i.e. Set 2, and not detected in at least 2 samples of their
617 respective contrast), we employed a single class test. If proteins were present in at least two
618 Twist and two control samples, we deemed this set suitable for a two-class comparison, i.e.
619 direct comparison of protein levels (set 3). The overall schema is represented in Fig. 1E and
620 summarised below:

621 Set 1: present preferentially in IPs (n ≥ 2)

622 Set 2: present preferentially in control IPs (n ≥ 2)

623 Set 3: present in Twist and control cell IPs (n ≥ 2)

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

628

629 **Protein Immunoprecipitation**

630 For the validation of protein interaction, hTWIST1 expressing MDCK cells were transfected
631 24 h before immunostaining using Lipofectamine® 3000 (Life Tech) according to
632 manufacturer instructions with one of the following plasmids: *pCMV-gfp-HA*, *pCMV-Tcf3-HA,*
633 *pCMV-Tcf4-HA, pCMV-Twist2-HA, pCMV-Ebf1-HA, pCMV-Ebf3-HA, pCMV-Sim2-HA,*
634 *pCMV-Tal1-HA.*

635 Cell pellets were thawed in 300 µL hypotonic lysis buffer (HEPES 20 mM, MgCl2 1 mM,
636 Glycerol 10%, Triton 0.5 %, DTT 1 mM, 1X Complete protease inhibitor [Roche], 1000 U/ml
637 Benzo nuclease) and incubated at room temperature for 15 min. Equal volume of
638 hypertonic lysis buffer (HEPES 20 mM, NaCl2 500 mM, MgCl2 1 mM, Glycerol 10%, DTT 1
639 mM, 1X Complete protease inhibitor) was applied to the lysate. Cells were further broken
640 down by passage through a gauge-25 needle 10 times and rotated at 4°C for 30 min before
641 to be centrifuged at 14,000 x g for 15 min. The supernatant was incubated with an α-TWIST1
642 or an α-FLAG antibody (2 ug/ ml) at 4°C for 2 h. 1/10 volume of Protein-G agarose beads
643 (Roche) was then added and the sample rotated for 30 min at room temperature. Beads

644 were finally washed with ice-cold wash buffer (1:1 mixture of the two lysis buffers) 6 times
645 and transferred to a new 1.5 ml tube before. Samples were eluted in 2x LDS loading buffer
646 (Life Technologies) at 70°C for 10 min and half of the eluate was loaded on SDS-PAGE in
647 parallel of the "input" control for western blot analysis.

648

649 **Western Blotting**

650 Proteins were extracted using RIPA lysis buffer (1X PBS, 1.5% Triton X-100, 1% IGEPAL, 0.5%
651 Sodium Deoxycholate, 0.1% SDS, 1 mM DTT, 1X Complete protease inhibitor [Roche]) for 30
652 minutes at 4°C while rotating. Lysed samples were then centrifuged at 15000 x g and the
653 supernatant was collected. Protein concentration was measured using the Direct Detect
654 spectrometer (Millipore). 20 µg of protein was denatured at 70°C for 10 min in 1X LDS
655 loading buffer. Protein electrophoresis and transfer was performed using the NuPage
656 system (Life Technologies, Cat. NP0322BOX), following manufacturer instructions.

657 Primary antibodies used were mouse monoclonal anti-TWIST1 (1:1000, abcam, Cat.
658 #ab50887), mouse anti-α-tubulin (1:1000, Sigma, Cat. #T6199), rabbit anti-HA (1:1000,
659 abcam, Cat. #ab9110), mouse anti-FLAG M2 (1:1000, Sigma, Cat. #F1804), mouse anti-TCF4
660 (1:500, Santa Cruz, Cat. #sc-166699), rabbit anti-TCF12 (1:400, Santa Cruz, Cat. # SC-357 1)
661 and mouse anti-GFP (1:1000, Thermo Fisher Scientific, Cat. # A-11120). Secondary
662 antibodies used were HRP-conjugated donkey anti-Rabbit IgG (1:8000, Jackson
663 Immunoresearch, Cat. #711-035-152) and HRP-conjugated donkey anti-Mouse IgG (1:8000,
664 Jackson Immunoresearch, Cat. #711-035-150).

665

666 **RNA-seq and data analysis**

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

669 RNA library prep (Truseq mRNA kit) and sequencing service (Hiseq2500, 100bpPE) were
670 performed by Macrogen (Seoul, Korea). Raw Fastq sequence files (20 Million reads per
671 sample) were imported into Galaxy (https://galaxy-mel.genome.edu.au/) and analysed
672 using a standard pipeline outlined in Fig. 3.8 (Anders *et al.*, 2013). FastQC package was used
673 to assess data quality and adapter sequences and low-quality reads were trimmed from the
674 original reads (Slidingwindow method in Trimmomatic package). A minimum of 50 bp read
675 length filter was applied. Alignment was performed using Tophat2 package using the built-in
676 mm10 genome. The resulting bam file was sorted and assessed visually before reads were
677 counted using the htseq-count package and Genecode mouse basic genome assembly
678 (GRCm38, vM10).

679 Read count values were imported into the R software for DEG analysis (Afgan *et al.*, 2016)).
680 Genes that were not expressed at a biologically meaningful level under any conditions were
681 discarded. A standard count per million (CPM) reads value threshold of 1 (or log-CPM value
682 of 0) was used for considering a gene to be expressed. To be kept for downstream analysis,
683 a gene had to be expressed in at least three samples across the entire experiment. After
684 filtering using this criteria, the number of genes was reduced to 14,165, approximately half
685 the original number. One replicate of T12 sample was removed from further analysis due to
686 low *Tcf12* induction level, and abnormal sample distance from other replicates in correlation
687 map. For EdgeR and limma-voom analysis, the data were normalised by library size before
688 further processing.

689 Pre-processed data were analysed with R/Bioconductor packageDESeq2, (Love *et al.*, 2014).
690 The common adjusted P-value (adjp) threshold of 0.05 and Log2FC of 2 was used. Two types
691 of gene-set analysis were applied: competitive gene-set analysis using clusterProfiler R
692 package (Yu *et al.*, 2012) in Figure 5G, and self-contained gene-set analysis using ROAST
693 package in Table S2 and summarised in Supplemental Fig. S4 (Wu *et al.*, 2010). In contrast to
694 commonly used competitive gene-set analysis, the self-contained test takes gene expression
695 into account and examines whether the genes in the set/pathway are differentially
696 expressed as a whole for different dimers. Motif analysis on DEGs was performed using the
697 Homer package (Heinz *et al.*, 2010).
698
699 **Plasmid construction**
700 Plasmids used in the BiFC experiments were a generous gift from Hu Lab and imported via
701 the Addgene depository. The plasmid backbone and the describing publication are listed in
702 Table S3 (Shyu *et al.*, 2006). *Twist1* and *Tcf12* coding sequences were cloned from cDNAs
703 generated from E9.5 mouse head embryos and inserted into the relevant expression
704 plasmids. Mouse homologous of S123A, R118H, T121E;S123E, T148A, A152P and I156N
705 variants were generated by site-directed mutagenesis as described previously (Liu and
706 Naismith, 2008) using primers listed in Supplemental Table S5.
707
708 **Bi-Fluorescent Complementation (BiFC) Experiments**
709 BiFC experiments and quantitative analysis were performed as described previously (Hu and
710 Kerppola, 2003). The day before transfection, C3H10T1/2 cells were seeded on 0.1%
711 gelatine coated coverslips at a density of 1.66x10^6 cells per 9.6 cm^2 (well of 6-well plate).
712 24 h later, transient transfection was performed using Lipofectamine 3000 (Life

713 Technologies) according to manufacturer's instructions. Equimolar quantities (1 μg each) of
714 plasmids encoding TWIST1-VN173, TCF12-CrN173 and wildtype or mutant TWIST1-CC155
715 (Fig. S1A) were co-transfected.

716 Cells were imaged 16-24 hours following transfection. Before imaging, cells were brought to
717 room temperature, washed twice with PBS 1X and fixed in 4% paraformaldehyde (dissolved
718 in PBS 1X) for 10 min. Coverslips were then washed three times with PBT (1% Tween 20 in
719 PBS 1X), once with 1% Trition X-100 in PBS 1X for 5 min and three times with PBT. This was
720 followed by incubation with 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich; 1:10000)
721 for 10 min at room temperature. Coverslips were then washed three times with PBT for 5
722 min in the dark and mounted using Fluoromount-G (Invitrogen).

723

724 All immunofluorescence slides were imaged using Zeiss Axio Imager A1 (Carl Zeiss,
725 Australia). TWIST1*-CC155/TCF12-CrN173 heterodimers were visualized using a CFP filter
726 set. A GFP filter set was used for detection of TWIST1*-CC155/ TWIST1-VN173 homodimers.
727 DAPI was imaged at 405 nm. For BiFC quantification, regions of the slides to be imaged were
728 randomly chosen and pictured at 20 X magnification. At least 1000 cells for each treatment
729 (based on DAPI counting) were sampled across three biological replicates. Fluorescence
730 intensity for the nucleus (DAPI stained) was quantified using ImageJ software and analysed
731 in R (scripts are available upon request).

732 After filtering for background fluorescence, using the same intensity threshold for all
733 treatment groups, the ratio between CFP/GFP was taken for each cell. Mean value of this
734 ratio was used to compare the CC155/CrN173 and CC155/VN173 dimer formation

735 efficiencies between wildtype or mutant TWIST1. Non-parametric Mann-Whitney U Test
736 was used for statistical analysis. TWIST1* represents mouse homologous of wildtype, S123A,
737 R118H, T121E; S123E, T148A, A152P or I156N variants.

738 **Web Resources**

739 OMIM: http://www.omim.org/

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

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

742 AQUARIA: http://aquaria.ws

Acknowledgements

Imaging analysis was performed at the ACRF Telomere Analysis Centre and proteomics analysis was performed at 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, the 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., N.F. 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, 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.

Liu, H. & Naismith, J. H. 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC biotechnology,* 8, 91.

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.

Murre, C. 2019. Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. *Genes & development,* 33, 6-25.

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.

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.

Soo, K., O'rourke, M. P., Khoo, P.-L. L., Steiner, K. A., Wong, N., Behringer, R. R. & Tam, P. P. 2002. Twist function is required for the morphogenesis of the cephalic neural tube and the differentiation of the cranial neural crest cells in the mouse embryo. *Developmental biology,* 247, 251-270.

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.

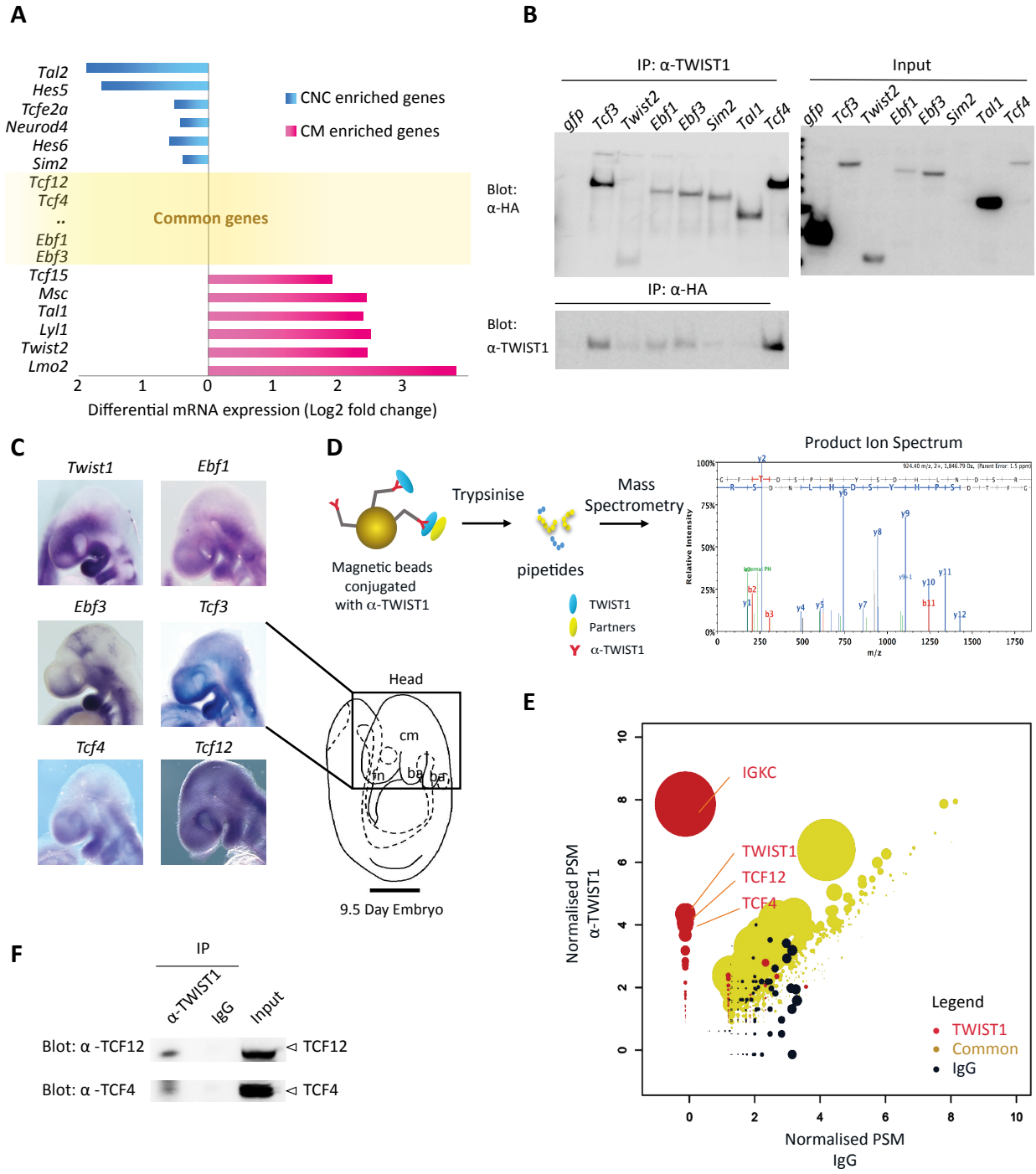
Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.-L., Visvader, J. E. & Smyth, G. K. 2010. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics,* 26, 2176-2182.

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.

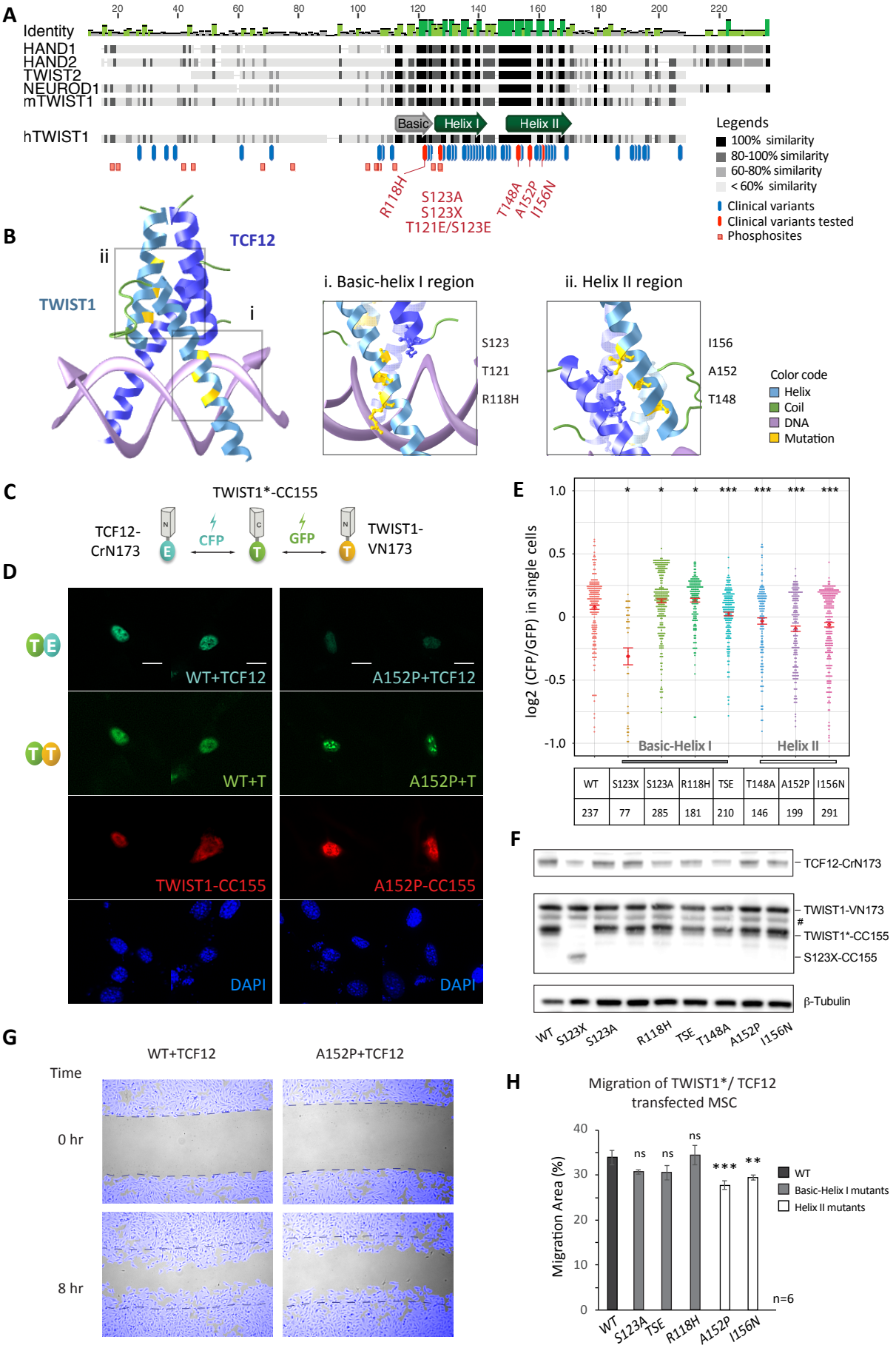
Xue, G. & Hemmings, B. A. 2013. PKB/Akt-dependent regulation of cell motility. *Journal of the National Cancer Institute,* 105, 393-404.

Yu, G., Wang, L.-G. G., Han, Y. & He, Q.-Y. Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology,* 16, 284-287.

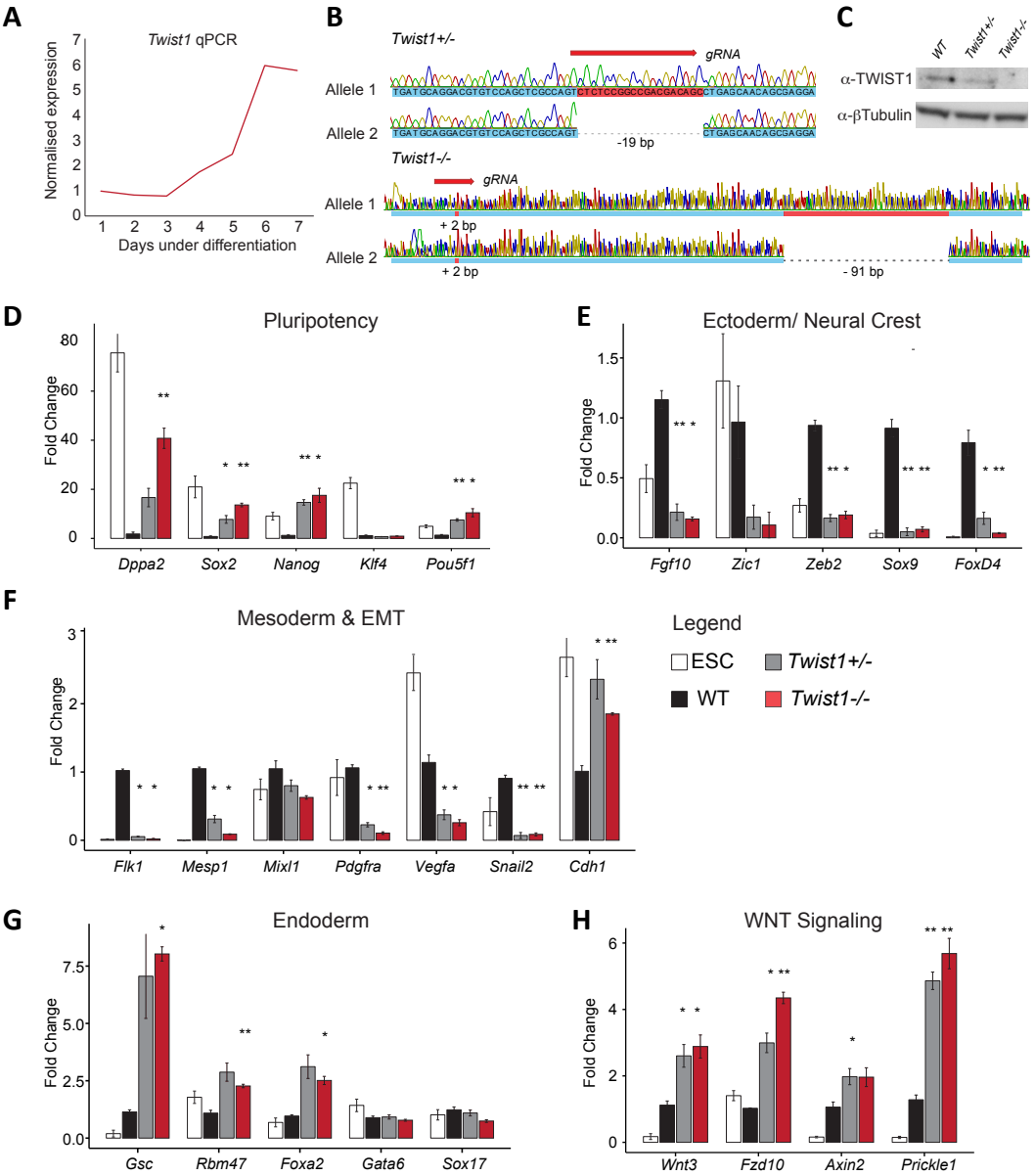
Fan_Fig1


A
Tal2
Hes5
Tcfe2a
Neurod4
Hes6
Sim2
Tcf12
Tcf4
..
Ebf1
Ebf3
Tcf15
Msc
Tal1
Lyl1
Twist2
Lmo2
CNC enriched genes
CM enriched genes
Common genes
2 1 0 1 2 3
Differential mRNA expression (Log2 fold change)
B
IP: α-TWIST1
Input
gfp Tcf3 Twist2 Ebf1 Ebf3 Sim2 Tal1 Tcf4
Blot: α-HA
IP: α-HA
Blot: α-TWIST1
C
Twist1
Ebf1
Ebf3
Tcf3
Tcf4
Tcf12
D
Magnetic beads conjugated with α-TWIST1
Trypsinise
pipetides
Mass Spectrometry
TWIST1
Partners
α-TWIST1
Product Ion Spectrum
Relative Intensity
m/z
Head
cm
fn
ba
9.5 Day Embryo
E
IGKC
TWIST1
TCF12
TCF4
Normalised PSM α-TWIST1
Normalised PSM IgG
Legend
TWIST1
Common
IgG
F
IP
α-TWIST1
IgG
Input
Blot: α -TCF12
TCF12
Blot: α -TCF4
TCF4

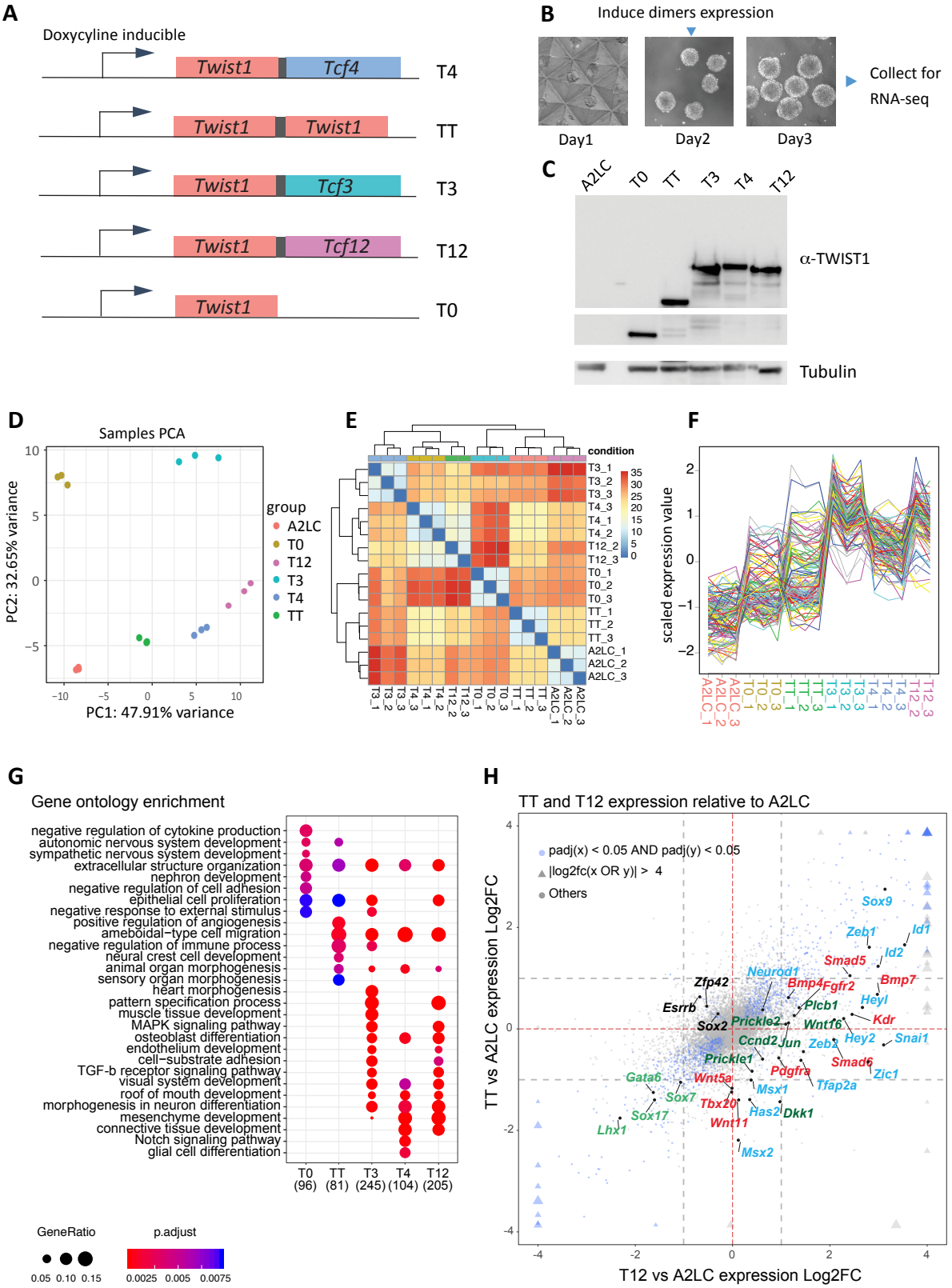
Fan_Fig2


A
Identity
HAND1
HAND2
TWIST2
NEUROD1
mTWIST1
hTWIST1
Basic
Helix I
Helix II
R118H
S123A
S123X
T121E/S123E
T148A
A152P
I156N
Legends
100% similarity
80-100% similarity
60-80% similarity
< 60% similarity
Clinical variants
Clinical variants tested
Phosphosites
B
TCF12
TWIST1
i. Basic-helix I region
ii. Helix II region
S123
T121
R118H
I156
A152
T148
Color code
Helix
Coil
DNA
Mutation
C
TWIST1*-CC155
TCF12-CrN173
CFP
GFP
TWIST1-VN173
D
WT+TCF12
A152P+TCF12
WT+T
A152P+T
TWIST1-CC155
A152P-CC155
DAPI
E
log2 (CFP/GFP) in single cells
Basic-Helix I
Helix II
WT S123X S123A R118H TSE T148A A152P I156N
237 77 285 181 210 146 199 291
F
TCF12-CrN173
TWIST1-VN173
TWIST1*-CC155
S123X-CC155
β-Tubulin
WT S123X S123A R118H TSE T148A A152P I156N
G
WT+TCF12
A152P+TCF12
Time
0 hr
8 hr
H
Migration of TWIST1*/ TCF12 transfected MSC
Migration Area (%)
WT
Basic-Helix I mutants
Helix II mutants
n=6
WT S123A TSE R118H A152P I156N

Fan_Fig3

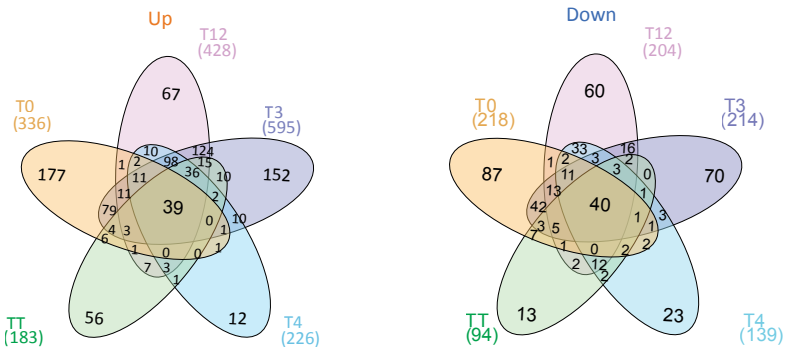

A
Twist1 qPCR
Normalised expression
Days under differentiation
B
Twist1+/-
gRNA
Allele 1
Allele 2
-19 bp
Twist1-/-
+ 2 bp
- 91 bp
C
WT
Twist1+/-
Twist1-/-
α-TWIST1
α-βTubulin
D
Pluripotency
Fold Change
Dppa2
Sox2
Nanog
Klf4
Pou5f1
E
Ectoderm/ Neural Crest
Fgf10
Zic1
Zeb2
Sox9
FoxD4
F
Mesoderm & EMT
Flk1
Mesp1
Mixl1
Pdgfra
Vegfa
Snail2
Cdh1
Legend
ESC
WT
Twist1+/-
Twist1-/-
G
Endoderm
Gsc
Rbm47
Foxa2
Gata6
Sox17
H
WNT Signaling
Wnt3
Fzd10
Axin2
Prickle1

Fan_Fig4

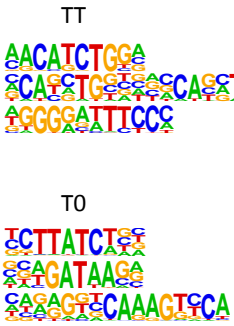

A
Doxycyline inducible
Twist1
Tcf4
T4
Twist1
Twist1
TT
Twist1
Tcf3
T3
Twist1
Tcf12
T12
Twist1
T0
B
Induce dimers expression
Day1
Day2
Day3
Collect for RNA-seq
C
A2LC
T0
TT
T3
T4
T12
α-TWIST1
Tubulin
D
Samples PCA
PC2: 32.65% variance
PC1: 47.91% variance
group
A2LC
T0
T12
T3
T4
TT
E
condition
F
scaled expression value
G
Gene ontology enrichment
negative regulation of cytokine production
autonomic nervous system development
sympathetic nervous system development
extracellular structure organization
nephron development
negative regulation of cell adhesion
epithelial cell proliferation
negative response to external stimulus
positive regulation of angiogenesis
ameboidal−type cell migration
negative regulation of immune process
neural crest cell development
animal organ morphogenesis
sensory organ morphogenesis
heart morphogenesis
pattern specification process
muscle tissue development
MAPK signaling pathway
osteoblast differentiation
endothelium development
cell−substrate adhesion
TGF-b receptor signaling pathway
visual system development
roof of mouth development
morphogenesis in neuron differentiation
mesenchyme development
connective tissue development
Notch signaling pathway
glial cell differentiation
T0 (96)
TT (81)
T3 (245)
T4 (104)
T12 (205)
GeneRatio
0.05 0.10 0.15
p.adjust
0.0025 0.005 0.0075
H
TT and T12 expression relative to A2LC
padj(x) < 0.05 AND padj(y) < 0.05
|log2fc(x OR y)| > 4
Others
TT vs A2LC expression Log2FC
T12 vs A2LC expression Log2FC
Sox9
Zeb1
Id1
Id2
Smad5
Neurod1
Bmp4
Fgfr2
Bmp7
Zfp42
Heyl
Esrrb
Plcb1
Kdr
Sox2
Prickle2
Wnt16
Hey2
Snai1
Ccnd2
Jun
Zeb2
Prickle1
Smad6
Pdgfra
Zic1
Gata6
Wnt5a
Tfap2a
Msx1
Sox7
Tbx20
Has2
Dkk1
Sox17
Wnt11
Lhx1
Msx2

Fan_Fig5

A



B

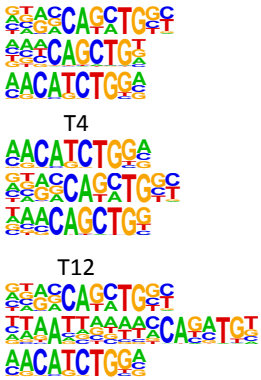


TT

Motif	P-value	# Target with motif	% Target with motif
ZBTB18(Zf)	1e-4	121	69%
Twist(bHLH)	1e-3	50	28%
NFkB-p65(RHD)	1e-3	122	69%

T0

Motif	P-value	# Target with motif	% Target with motif
Gata6(Zf)	1e-3	253	78%
Gata4(Zf)	1e-3	263	81%
HNF4a(NR)	1e-3	196	61%



T3

Motif	P-value	# Target with motif	% Target with motif
Atoh1(bHLH)	1e-9	513	90%
Ap4(bHLH)	1e-6	514	90%
ZBTB18(Zf)	1e-6	357	62%

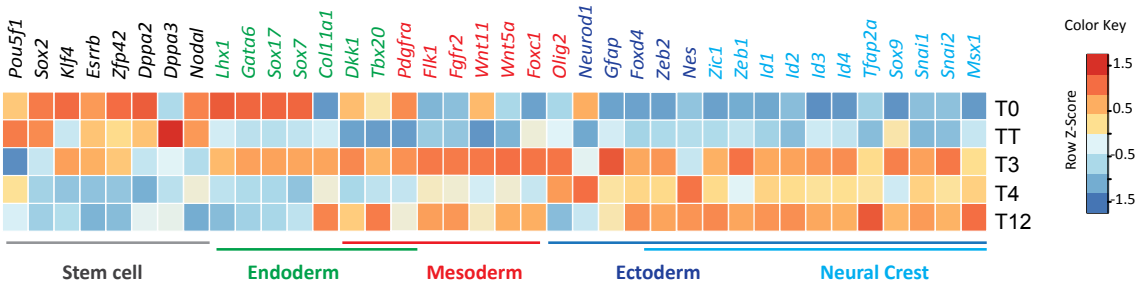
T4

Motif	P-value	# Target with motif	% Target with motif
ZBTB18(Zf)	1e-6	153	70%
Atoh1(bHLH)	1e-6	203	93%
Tcf21(bHLH)	1e-5	194	89%

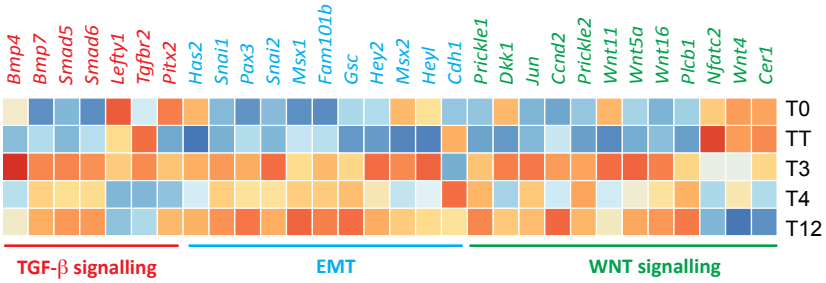
T12

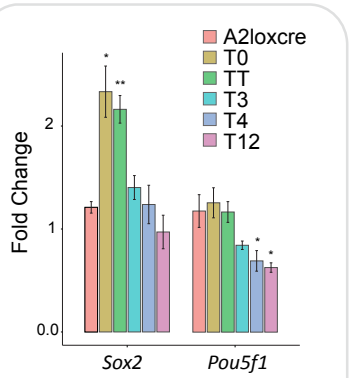
Motif	P-value	# Target with motif	% Target with motif
Atoh1(bHLH)	1e-9	374	91%
Pitx1:Ebox (Homeobox,bHLH)	1e-7	117	28%
ZBTB18(Zf)	1e-6	266	65%

C. Cell lineages

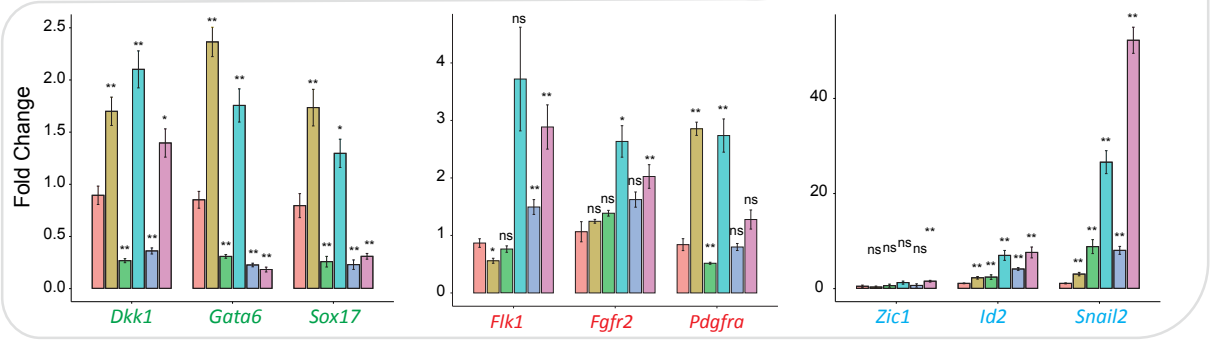


D. Pathways

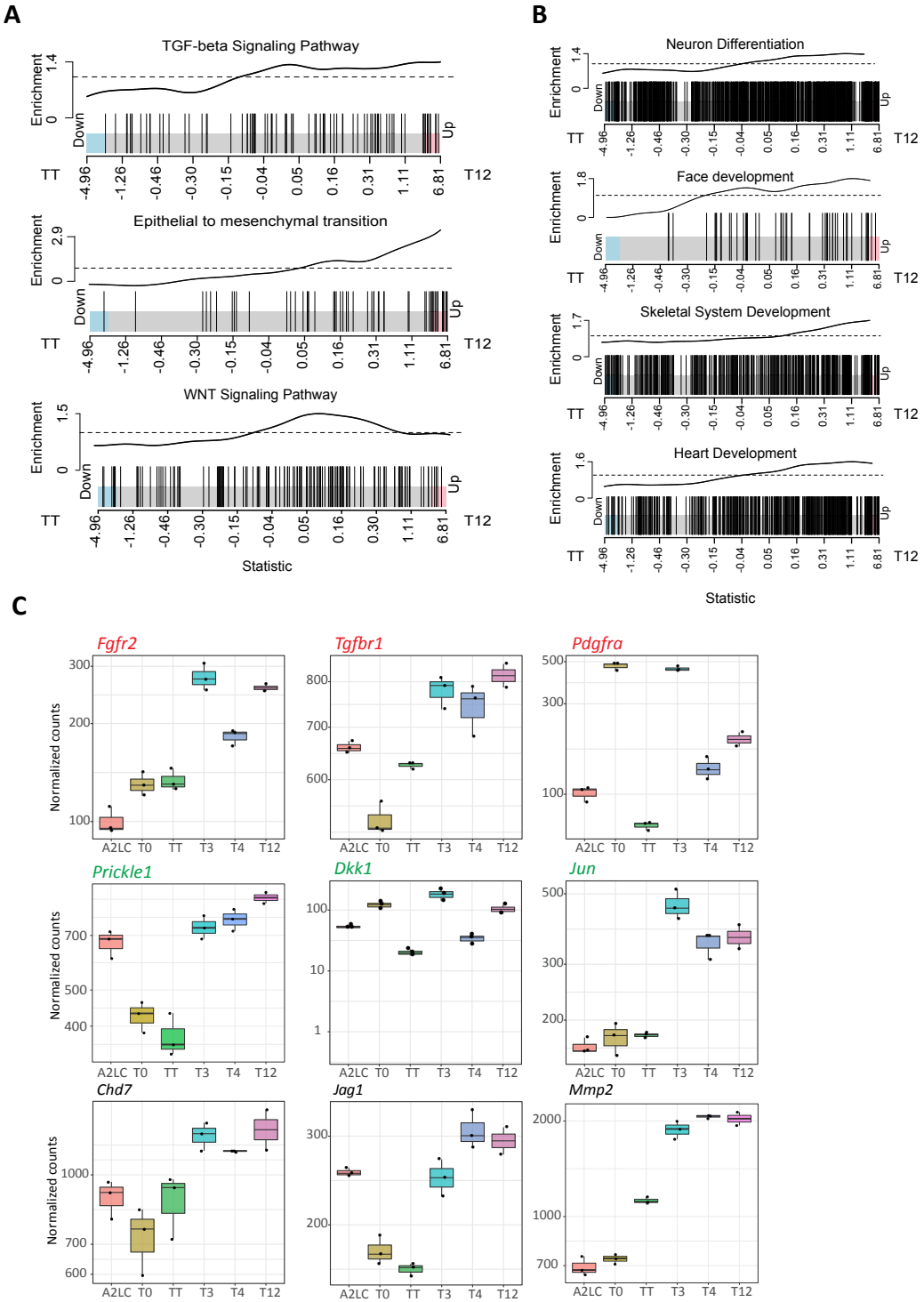




E



Fan_Fig6


A
TGF-beta Signaling Pathway
Epithelial to mesenchymal transition
WNT Signaling Pathway
Enrichment
Down
Up
TT
T12
Statistic
B
Neuron Differentiation
Face development
Skeletal System Development
Heart Development
C
Fgfr2
Tgfbr1
Pdgfra
Prickle1
Dkk1
Jun
Chd7
Jag1
Mmp2
Normalized counts
A2LC
T0
TT
T3
T4
T12