Cohesin composition and dosage independently affect early development in zebrafish

Authors

Labudina, A. A.,[1] Meier, M.,[1] Gimenez, G.,[1] Tatarakis, D.,[2,3] Ketharnathan, S.,[1,4] Mackie, B.,[1] Schilling, T. F.,[2] Antony, J.,[1] Horsfield J. A.[1]*

Affiliations

1. Department of Pathology, Dunedin School of Medicine, University of Otago, P.O. Box 913, Dunedin 9016, New Zealand.
2. Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697-2300, USA.
3. Present Address: Denali Therapeutics, 161 Oyster Point Blvd. South San Francisco, CA 94080, USA.
4. Present address: Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Rd, Ottawa, Ottawa, ON K1H 5B2, Canada.

* To whom correspondence should be addressed: julia.horsfield@otago.ac.nz

Abstract

Cohesin, a chromatin-associated protein complex with four core subunits (Smc1a, Smc3, Rad21 and either Stag1 or 2), has a central role in cell proliferation and gene expression in metazoans. Human developmental disorders termed "cohesinopathies" are characterised by germline mutations in cohesin or its regulators that do not entirely eliminate cohesin function. However, it is not clear if mutations in individual cohesin subunits have independent developmental consequences. Here we show that zebrafish *rad21* or *stag2* mutants independently influence embryonic tailbud development. Both mutants have altered mesoderm induction, but only homozygous or heterozygous *rad21* mutation affects cell cycle gene expression. *stag2* mutants have narrower notochords and reduced Wnt signaling in neuromesodermal progenitors as revealed by single cell RNA-sequencing.

Stimulation of Wnt signaling rescues transcription and morphology in *stag2*, but not *rad21* mutants. Our results suggest that mutations altering the quantity versus composition of cohesin have independent developmental consequences, with implications for the understanding and management of cohesinopathies.

Teaser

Viable zebrafish mutants show that cohesin complex quantity versus composition lead to different transcriptional and developmental outcomes in the early embryo.

MAIN TEXT

Introduction

Cohesin is a multiprotein ring-shaped complex that is highly conserved from yeast to humans. The vertebrate mitotic cohesin ring consists of two structural maintenance of chromosomes subunits, Smc1a, Smc3 and an α-kleisin subunit Rad21 [1, 2]. In vertebrates Rad21 interacts with either one of two Stromalin subunits, Stag1 or Stag2, and collectively these subunits are necessary for cohesin's association with DNA [3-5]. Additionally, Nipbl and Wapl modulate cohesin's residency on chromatin: Nipbl facilitates loading of cohesin onto DNA [6], while Wapl facilitates its release [7].

Cohesin is best known for its role in physically linking replicated sister chromatids to ensure the accurate transmission of genetic material to daughter cells during cell division [1]. In addition to mediating sister chromatin cohesion, the cohesin complex also functions to repair DNA double strand breaks [8-10]. Loss of functional cohesin results in mitotic arrest and cell death [11-13]. Only a small fraction of cohesin is necessary for sister chromatid cohesion [14] suggesting that the observed high levels of cohesin in certain non-dividing cell types has important non-cell cycle functions.

Cohesin also functions in three-dimensional (3D) genome organization and the regulation of gene expression [5, 15-19]. Loop extrusion activity by cohesin organizes DNA into topologically associated domains (TADs) that constrain the regulation of gene expression [20-24]. The CCCTC-binding factor, CTCF, acts as a barrier to cohesin and limits loop extrusion between convergent CTCF sites [25, 26], leading to the overlap of cohesin and CTCF at TAD boundaries [25, 27-29]. In addition, cohesin has gene regulatory functions that are independent of CTCF [30]. Sites bound by cohesin but not CTCF are frequent at tissue-specific enhancers and promoters [31]. Intra-TAD loops formed by cohesin can regulate transcription by mediating enhancer-promoter contacts [32, 33]. However, only a subset of enhancer-promoter contacts and DNA looping events appear to depend on cohesin [34-36].

Germline cohesin insufficiency gives rise to a spectrum of multifactorial developmental disorders collectively known as 'cohesinopathies' [4, 18]. Typically, cohesinopathies result from heterozygous mutations in cohesin subunits or their regulators [4]. Cohesinopathies are associated with developmental delay, a diverse range of developmental anomalies, and intellectual disability [37]. The best known cohesinopathy is Cornelia de Lange syndrome (CdLS, MIM #122470), a

multisystem disorder encompassing delayed growth, neurological and intellectual dysfunction, limb abnormalities and gastrointestinal defects [38-41]. Well over half of CdLS cases are caused by mutations in *NIPBL* [42, 43], with mutations in other cohesin-associated proteins accounting for a smaller subset of individuals with overlapping phenotypes. The specific presentation of CdLS varies according to the cohesin-associated protein affected by genetic changes [44, 45].

RAD21 (MIM #606462) is among the five extensively studied genes associated with CdLS [42, 46, 47]. Individuals with *RAD21* mutations display growth retardation, minor skeletal anomalies, and facial features that overlap with CdLS, but lack severe intellectual disabilities [48]. Mutations in *RAD21* are also linked with Mungan Syndrome (MIM #611376) [49], sclerocornea [50] and holoprosencephaly [51]. Most *RAD21* mutations associated with cohesinopathy are truncations, missense mutations or in-frame deletions that are predicted to interrupt the interaction between RAD21 and SMC1A, SMC3, or STAG1/2 [52, 53]. RAD21 physically bridges the SMC1A/SMC3 heads and facilitates the cohesin loading process, likely by controlling the amount that complexes with DNA [54]. Therefore RAD21 abundance has potential to directly modulate the quantity of cohesin complexes on DNA and its mutation or deficiency would result in reduction in cohesin dose. Interestingly, the RAD21 protein must be intact for stable cohesin binding and looping at CTCF-CTCF sites, and must be present but not necessarily intact for looped contacts inside of CTCF domains [55]. Further supporting evidence suggests that the cohesion and loop extrusion activities of cohesin can be separated experimentally and that cohesin uses distinct mechanisms to perform these two functions [56].

Individuals with STAG2 deficiency also display features of cohesinopathies [51, 57-60]. Loss-of-function mutations in *STAG2* on the X chromosome are associated with Mullegama-Klein-Martinez syndrome (MKMS, MIM #301022) in females but only missense mutations are tolerated in males [61]. Exome sequencing further established *STAG1* and *STAG2* variants in patients with cohesinopathy phenotypes as loss-of-function [62] and recently, loss-of-function variants of *STAG2* have been categorized as X-linked cohesinopathies with features of CdLS [58, 60]. For example an individual with a mosaic *STAG2* variant was described to have developmental delay, microcephaly, and hemihypotrophy of the right side [63]. A distinctive cohesinopathy involving Xq25 microduplication that exclusively affects *STAG2* gives rise to moderate intellectual disability, speech delay and facial dysmorphism [64]. Additionally, some cases exhibit structural brain malformations consistent with holoprosencephaly [51, 57, 58, 60]. Several molecular studies show that STAG1 and STAG2 paralogues have distinct roles in 3D genome organization, but overlapping roles in the cell cycle [16, 65-69]. Moreover, STAG

subunits can be detected at specific locations on DNA independently of the rest of the cohesin complex [67, 70]. Deficiency in STAG2 leads to the upregulation of STAG1 and the substitution of STAG1 for STAG2 in the cohesin complex such that *STAG2* mutation leads to altered cohesin composition [71, 72].

Dysregulated expression of multiple genes downstream of cohesin deficiency is thought to be the predominant cause of cohesinopathies [12, 73-76]. Because human cohesinopathies with different genetic causes present with diverse phenotypes, it is possible that cohesin subunits independently modulate the transcription function of cohesin during development. This idea has not yet been tested in the early embryo when the developmental changes in cohesinopathies are determined. In this study we compare the transcriptional and developmental consequences of depleting Rad21 with depletion of Stag2. Rad21 controls cohesin quantity on DNA [54] while Stag2 is thought to bind DNA independently and locate cohesin to enhancers [65, 67, 70]. Therefore, we expect *stag2* mutants to interfere with cohesin's gene expression functions without interfering with the cell cycle. Because *stag1b* and *stag2b* mutants are viable [77] and the effects of *rad21* deficiency are dose-dependent [78], zebrafish offer a unique opportunity to investigate how cohesin complex quantity, versus cohesin complex composition, affects cell fate decisions in the early embryo [79]. To explore this question, we focus on the tailbud as a stem cell model.

The tailbud, located at the posterior end of the developing embryo, contains two populations of bipotent stem cells known as neuromesodermal progenitors (NMPs) and midline progenitor cells (MPCs) [80, 81]. These cells continuously divide and differentiate into neuroectoderm and mesoderm by activating cell type-specific transcription. By analysing *rad21* heterozygous and homozygous mutants (reflecting cohesin dose) and *stag2b* mutants (reflecting cohesin type), we compare how the amount and composition of the cohesin complex affect transcription in tailbud cells. We find that although *rad21* heterozygous mutants are viable and fertile, they exhibit altered expression of thousands of genes in the tailbud including cell cycle regulators, demonstrating that decreased cohesin dose affects both cell cycle and gene expression. In contrast, cell cycle gene expression is largely unaffected in *stag2* homozygous mutants, which are also viable and fertile. However *stag2* mutants show a unique narrowing of the midline mesodermal domain that forms the notochord, which is restored by concomitant removal of *stag1*.

Therefore, although both *rad21* and *stag2* cohesin mutants show deficiencies in mesoderm derived from NMPs and MPCs, the underlying molecular mechanisms are remarkably dissimilar. Rad21 deficiency blocks NMP differentiation leading to lack of mesodermal derivatives, while loss of *stag2* mutants causes NMPs to downregulate Wnt signaling leading to epithelial to

mesenchyme (EMT) defects. Changes in phenotype and gene expression unique to *stag2* mutants are rescued by stimulation of Wnt signaling by GSK3 inhibition, to which Rad21-deficient embryos are impervious.

Results

Stag1b and Stag2b are the main functional Stags in zebrafish

Zebrafish have four Stag paralogues: Stag1a, Stag1b, Stag2a, and Stag2b. Individual *stag* mutant lines (except *stag2a*) were previously generated and are homozygous viable [77]. To determine which paralogs are crucial for zebrafish development, we analysed the consequences of combining *stag1a* and *stag2b* as well as *stag1b* and *stag2b* mutants.

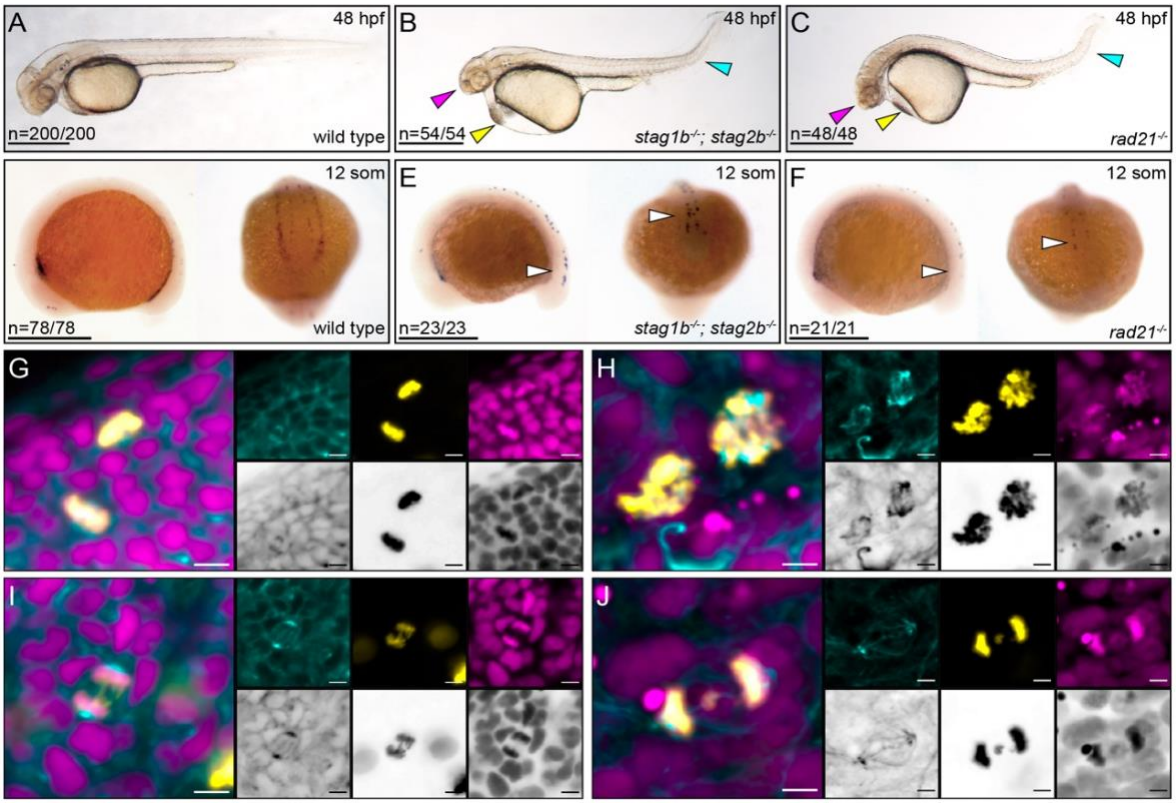

Fig. 1. Stag1b and Stag2b are the main functional Stag1 and Stag2 in Zebrafish. (A-C) Lateral views of representative wild type (A), *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ (B) and *rad21*$^{-/-}$ (C) embryos at 48 hours post-fertilization (hpf). Arrows indicate developmental anomalies: magenta for a small head, yellow for pericardial oedema, and cyan for a kinked tail. Scale bars are 500 µm. **(D-F)** Expression of *runx1* at 12 somites in wild type (D), *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ (E) and *rad21*$^{-/-}$ (F) embryos. Lateral and posterior views are shown. White arrows indicate the loss of *runx1* expression in PLM. Scale bars are 500 µm. The numbers in the lower left hand corner indicate the number of embryos with similar expression patterns. **(G-J)** Confocal images of cell cycle progression in wild type (G,I) and *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ (H,J) embryos at 48 hpf stained with anti-α-tubulin (cyan), anti-phH3 (yellow) antibodies and Hoechst (magenta). Images are maximum intensity projections of 3 (0.15 µm) optical sections taken from the tail region of 48 hpf embryos. Scale bars are 5 µm.

stag1b$^{-/-}$; *stag2b*$^{-/-}$ double mutant embryos were developmentally delayed compared to wild type, and by ~48 hours post-fertilization (hpf), mutant embryos had arrested in development presenting with small heads, pericardial oedema, upwards bending tails, and no blood circulation (Fig. 1B compared with A). This phenotype resembles *rad21*$^{-/-}$ mutant embryos, which die due to mitotic catastrophe (Fig. 1C) [12]. In contrast, *stag1a*$^{-/-}$; *stag2b*$^{-/-}$ embryos developed normally and are homozygous viable and fertile, although a small proportion (~5%) of *stag1a*$^{-/-}$; *stag2b*$^{-/-}$ embryos displayed hemorrhaging above the notochord at 48 hpf (Fig. S1A).

The gene encoding haematopoietic and neuronal transcription factor Runx1 is expressed in the anterior lateral plate mesoderm (ALM), the posterior lateral plate mesoderm (PLM), and in Rohon-Beard (RB) neurons in early zebrafish development (Fig. 1D) [82]. Rad21 is required for *runx1* expression in the PLM [12]. We previously found that *runx1* expression is normal in individual *stag* mutants [77]. However, we observed loss of *runx1* expression in the PLM of the *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ embryos and retained *runx1* expression in the ALM and RB neurons (Fig. 1E). This resembles changes in *runx1* expression in the *rad21*$^{-/-}$ mutant (Fig. 1F) [12], and is consistent with a requirement for an intact cohesin complex for *runx1* expression in the PLM. In contrast, *runx1* expression was normal in *stag1a*$^{-/-}$; *stag2b*$^{-/-}$ embryos (Fig. S1B).

We next examined the morphology of mitotic cells in *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ embryos at 48 hpf (Fig. 1G-J). In contrast to wild type embryos (Fig. 1G, I) condensed chromosomes were disorganized and abnormally distributed in *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ embryos, (Fig. 1H). Lagging chromosomes failed to properly separate during anaphase, resulting in some chromosomes remaining in cell centers (Fig. 1J). These findings suggest that cells in *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ mutants lack functional cohesin by 48 hpf, leading to a mitotic blockade. Individual *stag* mutants [77] as well as the *stag1a*$^{-/-}$*; stag2b*$^{-/-}$ double mutant, are homozygous viable. However, loss of both, *stag1b* and *stag2b*, is embryonic lethal and phenocopies the previously described *rad21*$^{-/-}$ mutant [12]. Based on these findings, we propose that Stag1b and Stag2b are the main functional Stag1 and Stag2 proteins in zebrafish.

Cell division proceeds normally in early stage cohesin mutant embryos

Loss of cohesin in *rad21*$^{-/-}$ homozygotes or *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ double mutants has different effects on *runx1* expression compared with viable mutations in *stag* genes. Therefore, we were curious to know whether cell cycle effects owing to cohesin deficiency could be responsible for gene expression changes, including *runx1*. Mitotic catastrophe occurs in embryos lacking

functional cohesin at 48 hpf (Fig. 1H,J) [12]. However, 16-somite *rad21*$^{-/-}$ homozygotes have sufficient maternally-deposited cohesin to continue growth for another 24 hours [12]. We chose to compare cell cycle progression in *stag2b*$^{-/-}$ homozygotes with *rad21*$^{-/-}$ homozygotes and heterozygotes to determine if these mutants alter cell cycle progression during early embryogenesis, the stage when *runx1* expression is disrupted.

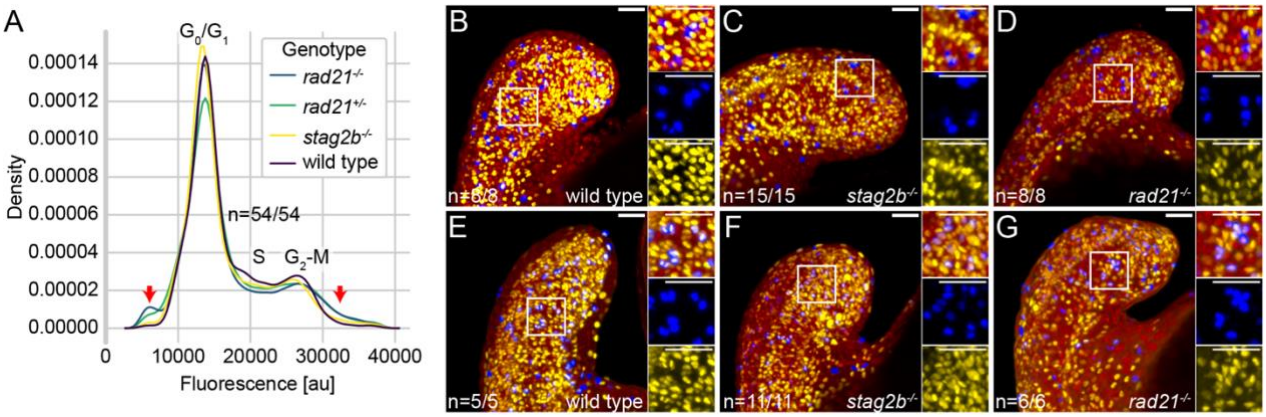

Fig. 2. The cell cycle is not blocked in cohesin mutants at the 16-somite stage. (A) Density plot (y-axis) showing the average signal of 3 replicates per genotype over fluorescence signal (DRAQ5, x-axis, artificial units). Red arrows indicate cells that are sub-G1, indicating potential chromosome loss, and cells that are >2n , indicating potential chromosome gain. **(B-G)** Confocal images showing S phases and M phases in wild type (B, E), *stag2b*$^{-/-}$ (C, F), and *rad21*$^{-/-}$ (D, G) tailbuds at ~16 hpf. S phases are detected with anti-BrdU (yellow) and M phases with anti-phH3 (blue) antibodies; nuclei are stained with Hoechst (red). BrdU incorporation was measured after incubation for 30 minutes (B-D) or 2 hours (E-G). Zoomed-in images of a selected area indicated by the box are shown. Images are maximum intensity projections of 33 (4.8 µm) optical sections. Scale bars are 40 µm.

Flow cytometry showed no significant differences in the proportions of cells in G1 (2n), S (2-4n), and G2/M (4n) phases between cohesin deficient tailbuds and wild type controls (Fig. 2A). However, in both homozygous and heterozygous *rad21*$^{-/-}$ (but not *stag2b*$^{-/-}$) embryos, we observed small populations of cells that were sub-G1 or >2n, possibly reflecting chromosome loss and gain. Using BrdU incorporation to mark S phases and phosphorylated histone H3 (phH$_3$) staining to mark G2/M cells, we found that S phase proceeds normally in cohesin mutants (Fig. 2B-G). Moreover, the presence of cells double-positive for BrdU and phH$_3$ indicated that cells progressed from S to M phase in cohesin mutant tailbuds (Fig. 2E-G). We conclude that cell division proceeds essentially normally in all cohesin deficient embryos at the tailbud stage, consistent with previous findings that even when cohesin complex quantity is substantially reduced, there remains sufficient cohesin to progress through the cell cycle during early embryogenesis [12, 78].

Cohesin complex quantity and composition affect tailbud gene transcription differently

The zebrafish embryonic tailbud contains neuromesodermal progenitors (NMPs) and midline progenitor cells (MPCs) as well as their neural and mesodermal derivatives (Fig. 3A) and therefore represents an ideal model to study changes in developmental gene transcription and cell fate decisions. Because Stag2 (rather than Stag1) is most likely to be involved in tissue-specific gene transcription [16, 65, 69, 83], we compared *stag2b* homozygous mutants with *rad21* homozygotes and heterozygotes to determine how the type of cohesin subunit mutation affects transcription in tailbuds. We performed bulk RNA-seq on 4 pools of 80 excised tailbuds from wild type, *rad21*$^{-/-}$, *rad21*$^{+/-}$, and *stag2b*$^{-/-}$ embryos stage-matched at 16 somites. Principal Component Analysis (PCA) separated samples into distinct groups based on their genotype (Fig. 3B). PC1 accounts for 61% of the variance and separated samples into two groups: homozygous and heterozygous *rad21* mutants vs wildtype and *stag2b*$^{-/-}$ mutants. PC2 accounts for an additional 11% of the variance and separated *rad21* homozygotes from heterozygotes, and *stag2b* mutants from wild type.

Normalized transcript counts of the cohesin subunits in the different genotypes showed that *rad21* mutation is associated with reduced transcript counts of the other cohesin core subunits, *smc1a* and *smc3*, and increased transcript counts of *stag* subunits. In contrast, transcription of core subunits was unaffected or increased in *stag2b* mutants, and *stag1b* transcript counts increased (Fig. S2). The findings are consistent with the idea that *rad21* mutation reduces cohesin quantity while *stag2b* mutation alters cohesin composition. Differential gene expression analysis revealed that 7250 genes are dysregulated in *rad21* homozygotes (Fig. 3C), 5144 in *rad21* heterozygotes (Fig. 3D) and 2054 in *stag2b* homozygotes (Fig. 3E). Notably, survivable changes in cohesin dose (*rad21*$^{+/-}$) and composition (*stag2b*$^{-/-}$) strongly affect transcription in the tailbud, indicating that normal levels and subunit makeup of the cohesin complex are important for gene expression.

Of the shared significantly dysregulated genes in cohesin mutant tailbuds, 311 were upregulated and 312 were downregulated in all cohesin deficient tailbuds (Fig. S3A,B), with the highest overlap between *rad21*$^{+/-}$ and *rad21*$^{-/-}$. Pathway enrichment analysis using Metascape showed that muscle organ development and energy metabolism were upregulated in all three genotypes, with the highest similarity between *rad21*$^{+/-}$ and *rad21*$^{-/-}$ (Fig. S3C). Of the downregulated gene pathways, none were conserved across all three genotypes, and more pathways were shared between *stag2b*$^{+/-}$ and *rad21*$^{-/-}$ than with *rad21*$^{+/-}$. A significant number of terms were unique to *rad21*$^{-/-}$ tailbuds including regulation of cell fate specification, suggesting

possible dysregulation of tailbud progenitor differentiation (Fig. S3D). Pathway enrichment analysis of significantly downregulated genes in *rad21*$^{-/-}$ tailbuds using Reactome revealed 26 significantly affected pathways, with top hits associated with mitosis and DNA damage repair (Fig. S4). The most affected pathway was cell cycle control, with 165 genes significantly downregulated in *rad21*$^{-/-}$. Although this pathway did not reach a significance threshold in other

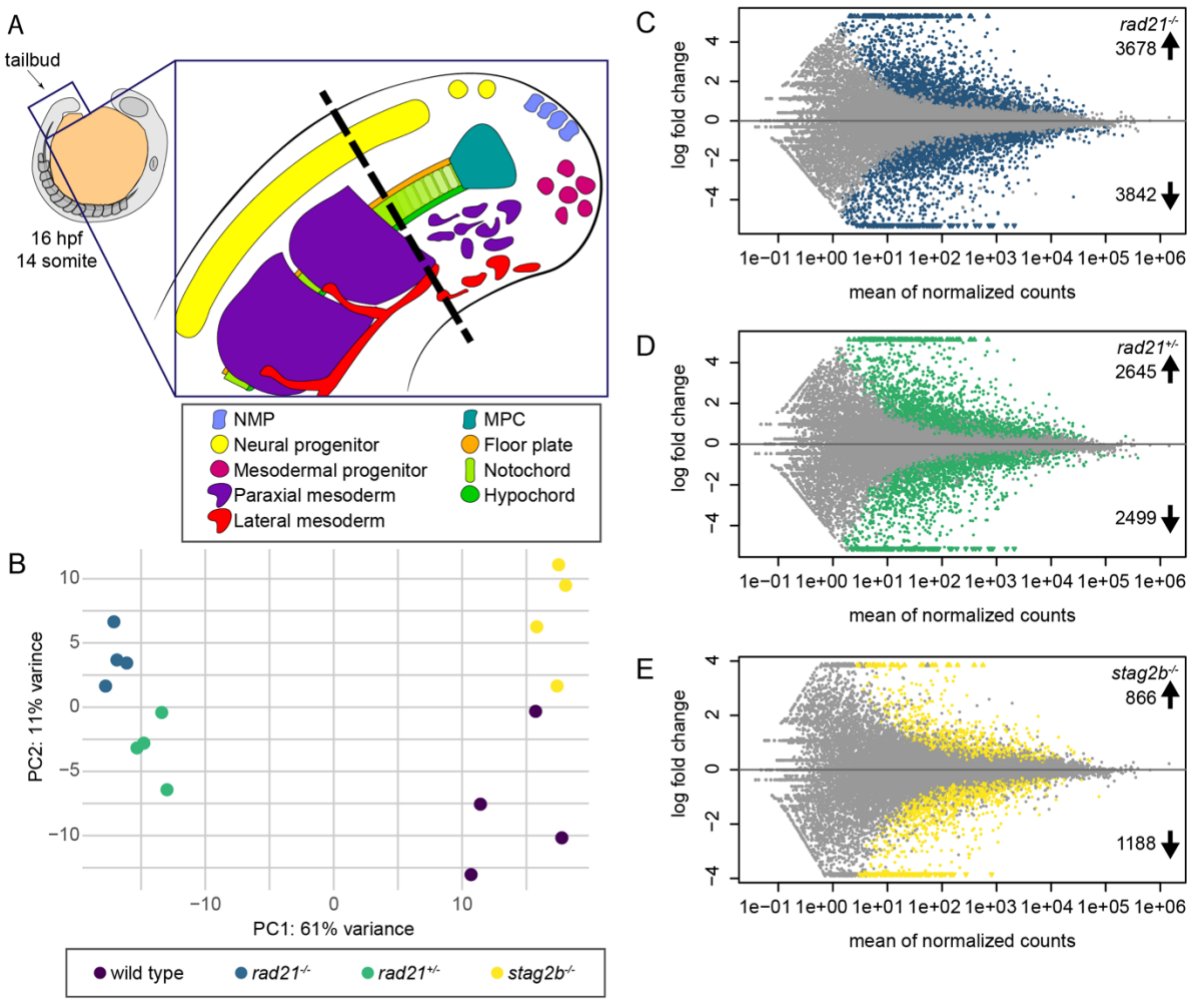

Fig. 3. Bulk RNA sequencing analyses of Rad21- and Stag2b-deficient tailbuds. (A) Schematic representation of progenitor cells and specialized tissues in the zebrafish tailbud. The zebrafish tailbud consists of two pools of bipotent progenitors: neuromesodermal progenitors (NMPs) and midline progenitor cells (MPCs). The dashed line shows the location of tailbud excision for RNA-seq. **(B)** Principal Component Analysis of gene expression in wild type and cohesin deficient tailbuds at the 16-somite stage. Genotypes are distinguished by colour: wild type samples are displayed in purple, *rad21*$^{-/-}$ in blue, *rad21*$^{+/-}$ in green, and *stag2b*$^{-/-}$ in yellow. **(C-E)** The MA (M (log ratio) and A (mean average) scales) plots display changes in gene expression in *rad21*$^{-/-}$ (C), *rad21*$^{+/-}$ (D), and *stag2b*$^{-/-}$ (E) compared to the wild type tailbuds. Each dot represents a gene, with colored dots indicating those with significant (5% false discovery rate, FDR) changes in expression. 7520 genes were dysregulated in *rad21*$^{-/-}$ tailbuds (3678 up- and 3842 downregulated), while 5144 genes were dysregulated in *rad21*$^{+/-}$ tailbuds (2645 up- and 2499 downregulated). In contrast, *stag2b*$^{-/-}$ tailbuds had substantially fewer dysregulated genes (2054: 866 up- and 1188 downregulated).

mutants, 79 cell cycle genes were significantly downregulated in *rad21*$^{+/-}$ and 13 in *stag2b*$^{-/-}$ (Data S1-3).

The number of shared dysregulated genes between genotypes suggests that transcriptional changes in *rad21*$^{+/-}$ mutants more closely resemble *rad21*$^{-/-}$ than *stag2b*$^{-/-}$ mutants (Fig. 3B). Transcriptional changes reflect genotype rather than viability through to adulthood: *rad21*$^{+/-}$ and *stag2b*$^{-/-}$ mutants are viable and *rad21*$^{-/-}$ mutants are not. Additionally, the results suggest that despite normal cell cycle progression in *rad21* mutants during early embryogenesis (Fig. 2), strong transcriptional changes relate to the expression of cell cycle genes in this genotype.

Subunit-specific effects of cohesin deficiency on transcription in tailbud cell populations

To assess how cohesin deficiency versus composition affects cell fate decisions in the tailbud we used the bulk RNA-seq data to quantify the expression of genes that mark progenitor cells and their derivatives. We used *rad21*$^{-/-}$ as a genotype that represents cohesin deficiency and *stag2b*$^{-/-}$ as a genotype that corresponds to altered cohesin composition.

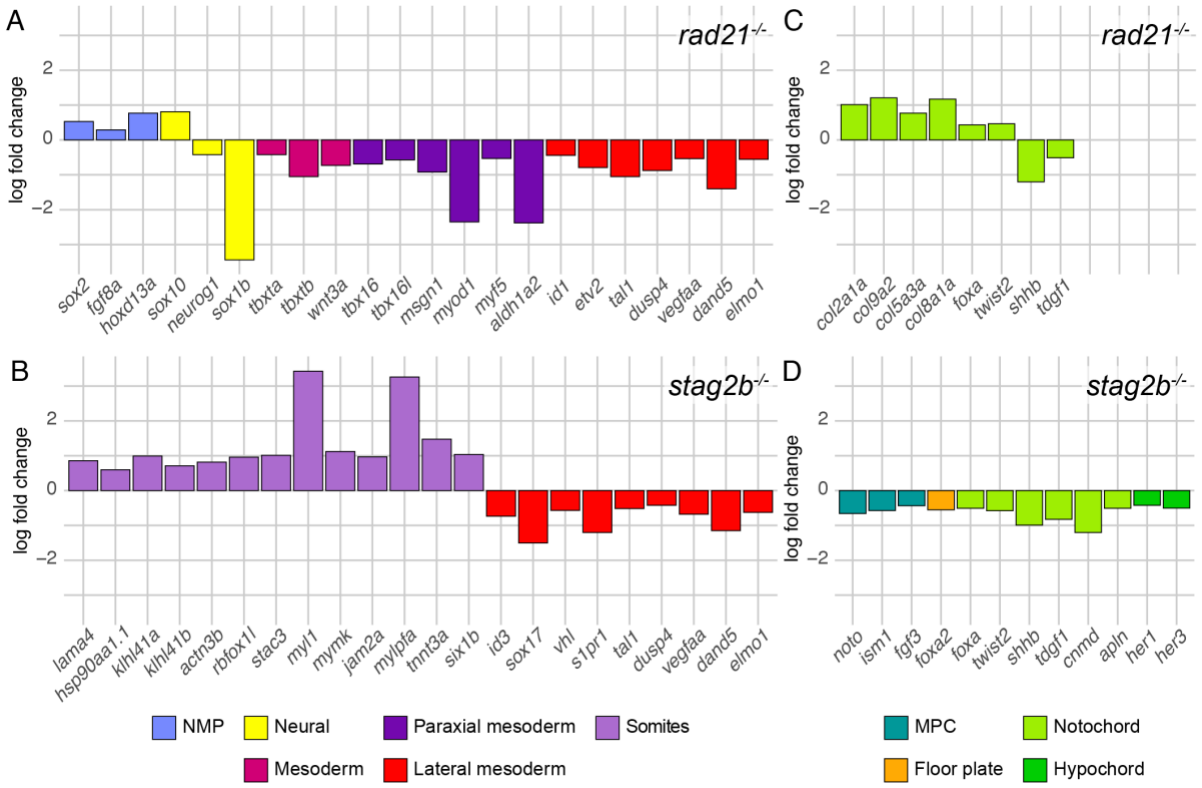

Fig. 4. Expression of genes that mark progenitor cells and their derivatives in *rad21* and *stag2b* homozygous mutant tailbuds. (A-D) The bar graphs display log2 fold changes significantly (5% FDR) dysregulated marker genes in *rad21*$^{-/-}$ (A, C) and *stag2b*$^{-/-}$ (B, D) tailbuds compared to wild type. The different categories of marker genes are represented by different colors as specified in the key.

Tailbud NMPs give rise to mesoderm and neuronal fates, while MPCs give rise to floorplate, notochord and hypochord (Fig. 3A). In *rad21*$^{-/-}$ tailbuds, genes marking NMPs were upregulated, neural genes were dysregulated (both up- and downregulated) and genes marking all mesoderm fates were downregulated, including lateral mesoderm that may not derive from NMPs (Fig. 4A). In contrast, NMP and mesoderm marker genes were more subtly affected in *stag2b*$^{-/-}$ tailbuds, with non-significant downregulation of *tbxta*, *tbxtb*, *tbx16* and *msgn1* (see Fig. S5 for additional expression data). *stag2b*$^{-/-}$ tailbuds had increased expression of genes that mark mature somites and decreased expression of genes marking lateral mesoderm (Fig. 4B).

Genes that mark MPCs were expressed normally in *rad21*$^{-/-}$. However, genes encoding notochord-specific collagens were upregulated, and some notochord markers were significantly dysregulated (Fig. 4C). In contrast, genes expressed in MPCs and midline tissues derived from MPCs were significantly downregulated in *stag2b*$^{-/-}$ tailbuds (Fig. 4D). The results suggest that *rad21* deficiency causes a block in NMP differentiation, while *stag2b* mutation either affects the composition of mesoderm, or mesoderm gene expression, in tailbuds. Moreover, *rad21* mutation had little effect on midline progenitors (with some effect on MPC derivatives) while *stag2b* mutation reduced transcription of genes expressed in MPCs and all derivatives.

***rad21* and *stag2b* mutants have different tailbud phenotypes**

We next investigated if gene expression changes reflect gross developmental changes in the tailbud in *rad21* and *stag2b* mutants by imaging tailbud cell populations. NMPs are marked by *sox2* and *tbxta* co-expression (Fig. S6). MPCs, a thin band of cells at the end of the notochord, also co-express *sox2* and *tbxta*. Mesoderm progenitors express *tbxta* but not *sox2*, and differentiate into paraxial mesoderm, labelled by *tbx16* expression. *sox2* expression alone labels neural progenitors, lateral mesoderm, the floor plate, and the hypochord (Fig. 3A; Fig. S6) [81, 84]. We used HCR RNA-FISH to visualize the distribution of *sox2*, *tbxta*, and *tbx16* transcripts in cohesin deficient tailbuds (Fig. 5).

In *rad21*$^{-/-}$ homozygotes (Fig. 5B compared with A) and in *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ double mutants (Fig. 5F compared with D) *sox2* expression was expanded at the posterior wall of the tailbud, and the zone of *sox2* expression extended into mesoderm progenitors accompanied by a reduction of *tbxta* expression in these cells. Expression of *tbx16* was restricted to a smaller area than in wild type. Approximately two-thirds of heterozygous *rad21*$^{+/-}$ embryos displayed similar expression changes, resembling homozygotes (Fig. 5C). Like *rad21* mutants, *stag2b*$^{-/-}$ mutants had expanded *sox2* expression in the posterior wall of the tailbud (Fig. 5E compared with D).

However, in *stag2b*$^{-/-}$, *tbx16* expression appeared normal, while the notochord, visualized by *tbxta* expression, was narrower and did not widen at the posterior end where MPCs reside. Ectopic expression of *sox2* was also observed in this region (Fig. 5E).

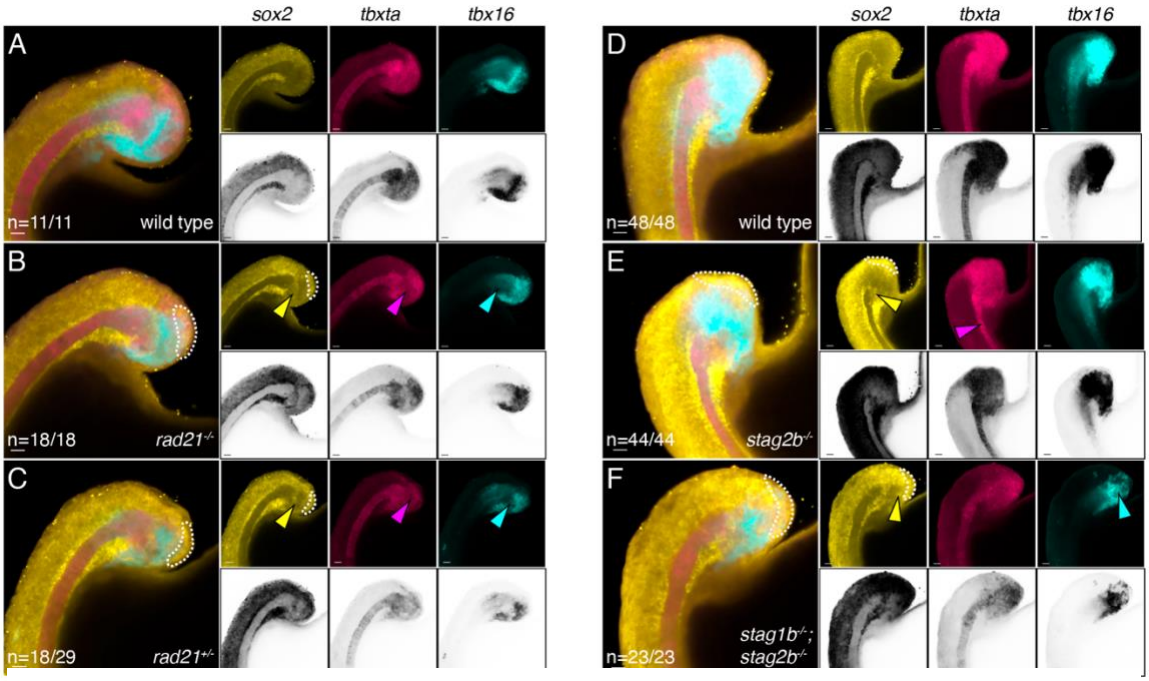

Fig. 5. Distribution of *sox2*, *tbxta*, and *tbx16* transcripts in cohesin-deficient tailbuds. (A, D) wild type, **(B)** *rad21*$^{-/-}$, **(C)** *rad21*$^{+/-}$, **(E)** *stag2b*$^{-/-}$, and **(F)** *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ zebrafish tailbuds at the 16-somite stage showing expression of *sox2* (yellow), *tbxta* (magenta), and *tbx16* (cyan). Increased *sox2* expression in the NMP region is outlined with a dashed line, yellow arrows indicate ectopic expression of *sox2* and pink arrows point to the loss of *tbxta* expression in the region of mesodermal induction (B, C) and the narrow notochord (E), while cyan arrows indicate a decrease in *tbx16* expression. Images are maximum intensity projections of 3 (4.8 µm) optical sections. Scale bars are 20 µm. The number of embryos with each expression pattern out of the total analyzed is noted at the bottom left of the merged panels.

We measured the thickness of the notochord (as defined by *tbxta* expression) in wild-type and cohesin-deficient embryos (Fig. 6A,B) and confirmed that notochords were significantly narrower ($p \leq 0.0001$) in *stag2b*$^{-/-}$ embryos (Fig. 6C). In contrast, notochord width in *rad21* homozygous and heterozygous embryos was similar to wild type (Fig. 6C). Surprisingly, *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ double mutant embryos had notochords that were normal width (Fig. 5F; Fig. 6C). Therefore, the narrow notochord phenotype was unique to *stag2b*$^{-/-}$, suggesting that the loss of Stag2b impacts MPC differentiation. Strikingly, this phenotype is "rescued" in *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ mutants, suggesting that complete cohesin loss is epistatic to the narrowed notochord in *stag2b*$^{-/-}$ mutants.

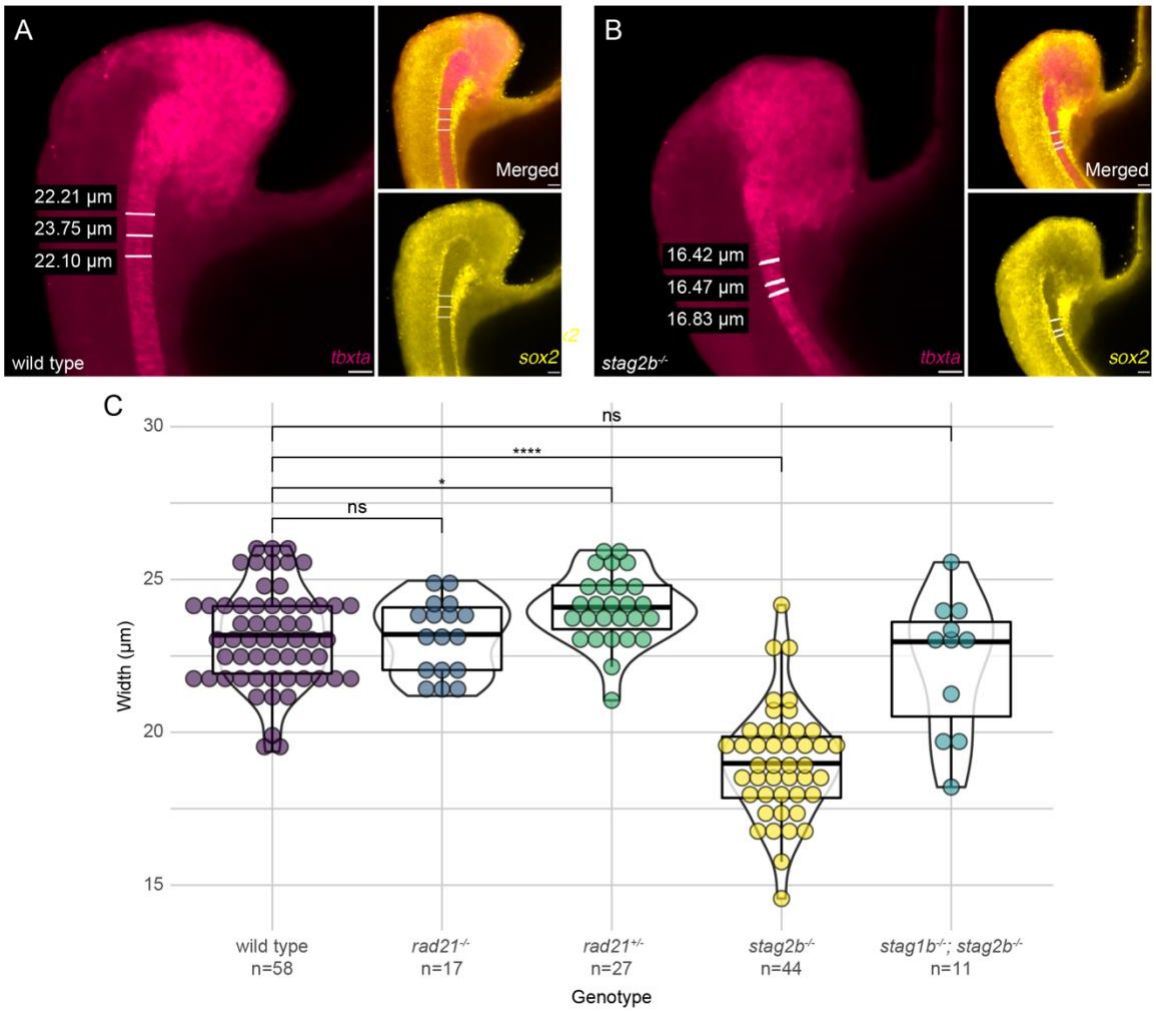

Fig. 6. Narrower notochords in *stag2b* mutants are rescued by additional *stag1b* mutation. (A, B) Example of notochord width measurement using *tbxta* expression and absence of *sox2* expression. The scale bar is 20 µm. **(C)** Violin plots with overlaid box plots visualizing measurements of notochord width. The genotype and the number of embryos measured in each group are indicated on the x-axis. Significance was determined using an unpaired t-test: * $p < 0.05$, **** $p < 0.0001$.

Altogether, our results indicate that different cohesin mutations have different effects on cell populations in the tailbud. Loss of cohesin quantity in *rad21* mutants and *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ double mutants caused reduction of *tbx16* and expansion of *sox2* expression, consistent with lack of mesoderm induction. In contrast, *stag2b* mutation (which alters cohesin composition) leads to a narrower notochord.

Altered cell populations in *stag2b*$^{-/-}$ tailbuds likely result from downregulated Wnt signaling in NMPs

Transcription dysregulation in homozygous and heterozygous *rad21* tailbuds is strongly associated with the cell cycle and developmental changes in the tailbud indicate a block in NMP differentiation. In contrast, transcription dysregulation in *stag2b*$^{-/-}$ tailbuds has no cell cycle association yet these mutants appear to have altered mesoderm and notochord gene expression together with physical changes to the notochord. It is possible that altered cohesin composition

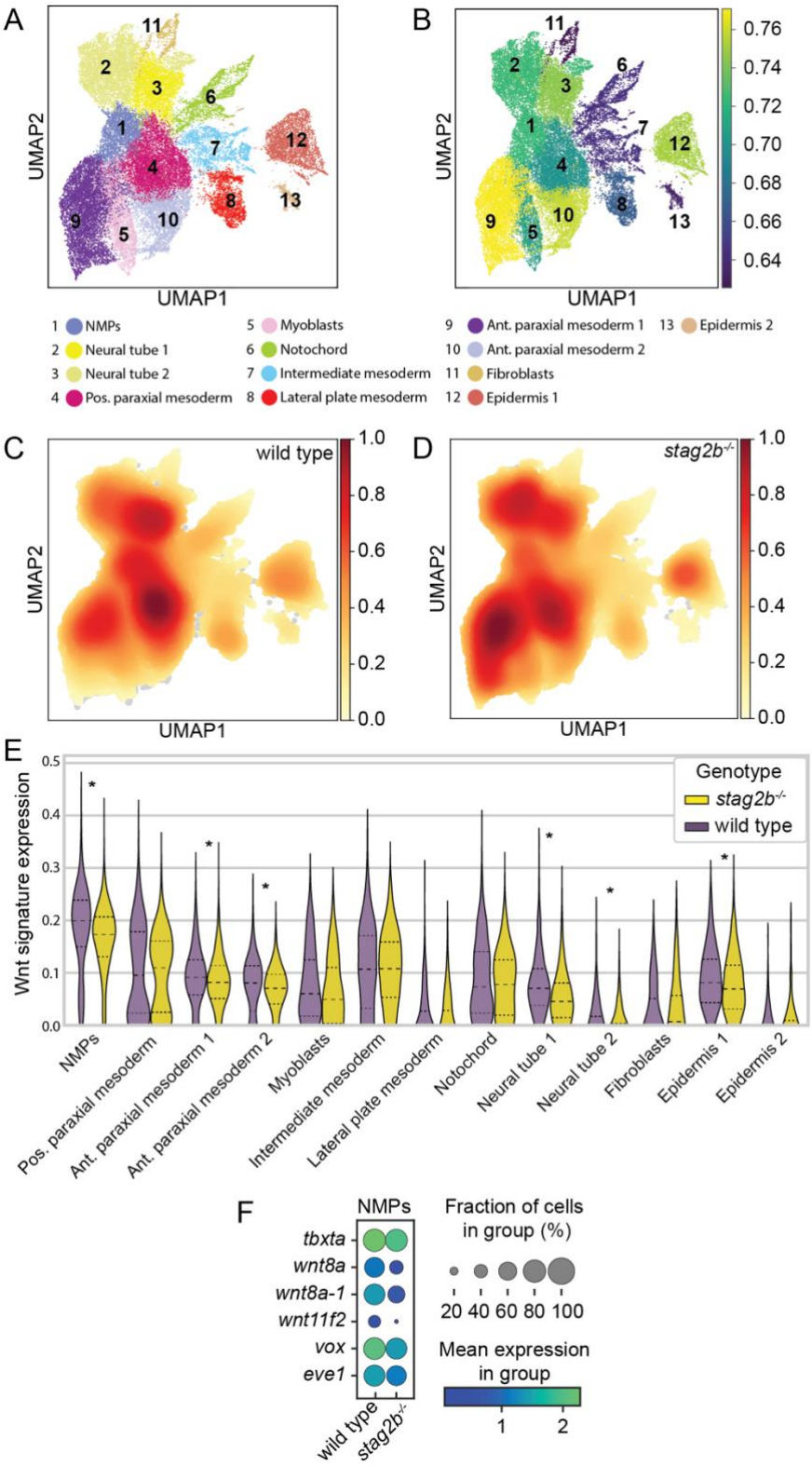

Fig. 7. Single cell RNA-seq of tailbuds from embryos at the 16-somite stage shows disruption of Wnt signaling in *stag2b*$^{-/-}$ NMPs. (A) Uniform manifold approximation and projection (UMAP) dimensional reduction of two integrated datasets of wild type (15,298 cells) and *stag2b*$^{-/-}$ (21,278 cells) tailbud samples (total 36,576 cells) with clustering of the major cell types. **(B)** UMAP representation of gene expression differences in *stag2b*$^{-/-}$ tailbuds calculated using Augur area under the receiver operating characteristic curve (AUC). The AUC ranges from 0 to 1, where values closer to 0.7 denote more minor differences in the *stag2b*$^{-/-}$ genotype (purple shades), and values closer to 1 denote larger changes in the *stag2b*$^{-/-}$ genotype (yellow shades). Visualization of the average cell density within wild type **(C)** and *stag2b*$^{-/-}$ **(D)**, using embedding density. Darker colors (red) correspond to denser regions. **(E)** Violin plot of Wnt gene expression signature (log-normalized) among different cell types in *stag2b*$^{-/-}$ (yellow) and wild type (purple) embryos. Wilcoxon rank-sum test with 5% FDR. **(F)** Dot plot showing the mean expression of differentially expressed Wnt ligands in the NMP cluster in wild type and *stag2b*$^{-/-}$ embryos.

through *stag2b* mutation has unique, cell cycle-independent effects on cell fate in the tailbud. We chose to investigate this possibility further using single-cell RNA-sequencing of *stag2b*$^{-/-}$ tailbuds compared with wild type at the 16-somite stage.

We integrated the single-cell RNA-seq datasets from wild type and *stag2b*$^{-/-}$ tailbud and annotated clusters representing major cell types based on their gene expression profiles (Fig. S7A). All annotated clusters were present in both wild type and *stag2b*$^{-/-}$ samples (Fig. 7A), although in slightly different proportions (Fig. S7B). The biggest changes in cell frequencies in *stag2b*$^{-/-}$ tailbuds compared to the wild type were evident in the anterior paraxial mesoderm 1 cluster (Fig. S7B). Although changes in other clusters were minor, they did reflect the variation in expression of cell type-specific markers observed in the bulk RNA-seq analysis (Fig. 4C,D).

Subsequently, we examined the gene expression differences within each cluster between the two genotypes (Fig. 7B). The clusters representing anterior paraxial mesoderm 1 and 2 displayed the most significant differences in gene expression between wild type and *stag2b*$^{-/-}$ samples, whereas the notochord, intermediate mesoderm, fibroblast and epidermis 2 clusters exhibited the least difference in gene expression profiles. Additionally, the cell density UMAPs indicated increased densities in anterior paraxial mesoderm 1 cluster (Fig. 7C, D). Collectively, these results suggest that the biggest changes in cell proportion and gene expression within each cell occur within the paraxial mesoderm lineage in the *stag2b*$^{-/-}$ tailbuds.

Given that midline mesoderm forms from NMPs, we analyzed the top 25 differentially expressed genes in this cluster (Fig. S8, Data S4) with *wnt8a* being among the most downregulated in *stag2b*$^{-/-}$. The important role of Wnt signaling in NMP differentiation and mesoderm formation prompted us to explore the Wnt gene expression signatures (Data S5) across various cell types. We observed significant downregulation of Wnt signatures in the NMPs and anterior paraxial mesoderm clusters (Fig. 7E, Data S6). *tbxta* and Wnt ligands (*wnt8a*, *wnt8a-1* and *wnt11f2*) were downregulated in the NMPs, as well as Wnt responsive genes *vox* and *eve1* (Fig. 7E-F). In summary, our results suggest that mesoderm formation from the NMPs is dysregulated in *stag2b*$^{-/-}$ tailbuds and these observed changes may be due to downregulation of Wnt signaling in these cell types.

Wnt stimulation rescues transcription in *stag2b* but not *rad21* mutant tailbuds

We next determined whether Wnt stimulation could restore transcription in cohesin deficient tailbuds. We performed RNA-seq on tailbuds of embryos treated from shield stage to 16-somite stage with the Wnt agonist, BIO (6-bromoindirubin-3’-oxime), which is a GSK3

inhibitor. Subsequently, we conducted interaction analysis (combined effect of genotype and

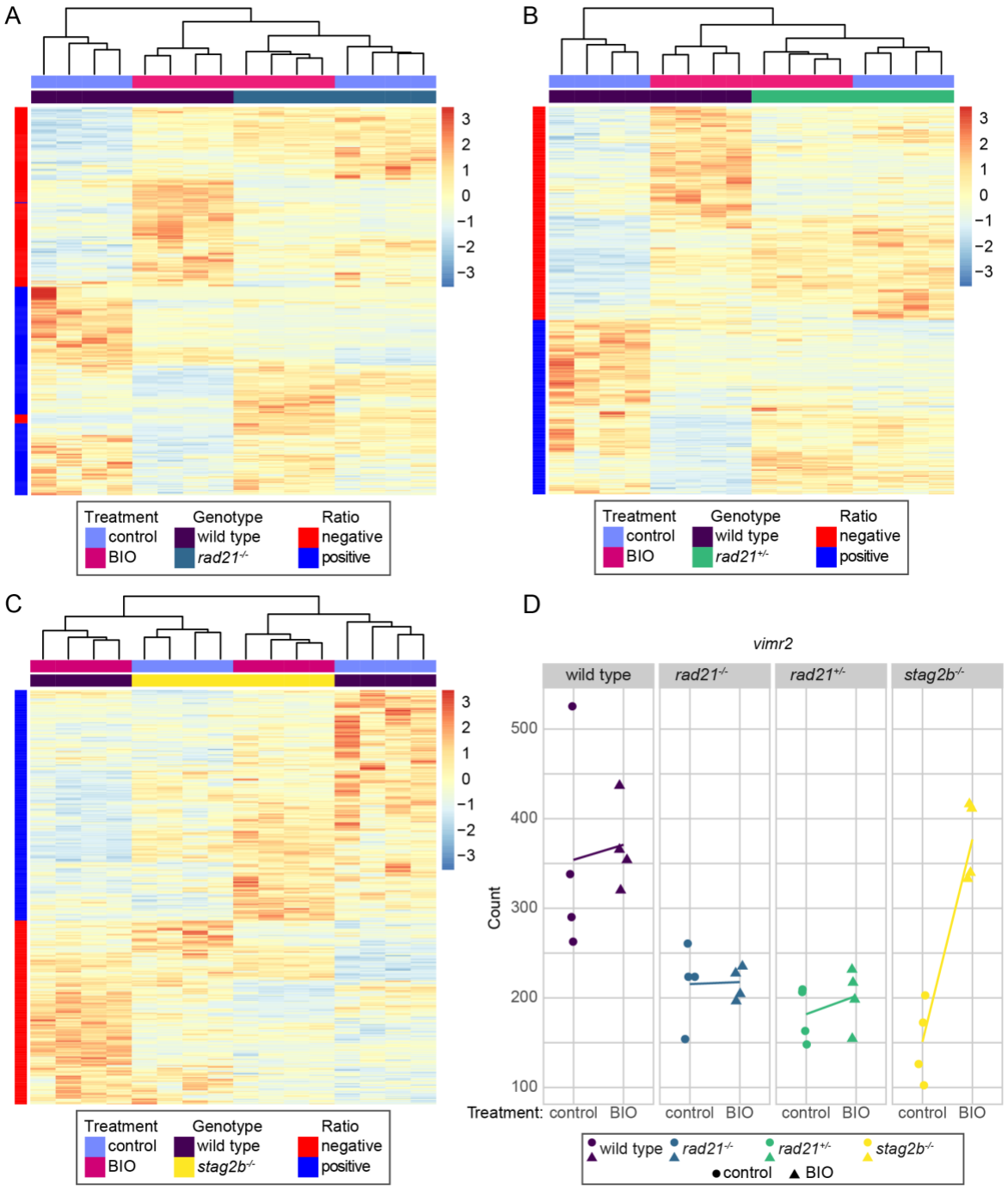

Fig. 8. Wnt stimulation normalizes gene expression in *stag2b*$^{-/-}$ but not in *rad21*$^{-/-}$ or *rad21*$^{+/-}$ tailbuds. Embryos were treated from shield stage with 2.5 µM BIO, then tailbuds were collected at 16 somites. 4 replicate pools of 80 tailbuds were used per condition for RNA-seq. The heatmaps display expression levels of the genes that responded differently to BIO stimulation in cohesin mutant genotypes compared with wild type as determined by an interaction analysis. Heatmaps display results from 4 replicates of **(A)** *rad21*$^{-/-}$, **(B)** *rad21*$^{+/-}$, and **(C)** *stag2b*$^{-/-}$ versus wild type. Red and blue colors indicate upregulation and downregulation, respectively, compared to the mean expression. **(D)**, *vimr2* expression is rescued by BIO stimulation in *stag2b*$^{-/-}$ but not in *rad21*$^{-/-}$ or *rad21*$^{+/-}$. Dot plots illustrate the transcript counts of *vimr2* in wild type (purple), *rad21*$^{-/-}$ (blue), *rad21*$^{+/-}$ (green), and *stag2b*$^{-/-}$ (yellow). The x-axis indicates the treatment status, and the y-axis represents the normalised counts. Lines connect the means of the counts for each sample group.

treatment) to identify genes exhibiting differential responses to Wnt stimulation in cohesin deficient tailbuds compared to wild type. Heatmaps were used to display clustering of the differentially responsive genes (Fig. 8A-C).

In *rad21*$^{-/-}$ and *rad21*$^{+/-}$, the genotype had a stronger effect on clustering of differentially responsive genes than BIO treatment. Genes identified as responding differently to BIO treatment in *rad21* mutants compared to wild type (395 in *rad21*$^{-/-}$ and 467 in *rad21*$^{+/-}$) cluster together in the dendrograms regardless of BIO treatment (Fig 8. A, B). Primarily, expression of these genes differs from wild type by being strongly responsive to BIO in wild type, and much less responsive to BIO with homozygous or heterozygous *rad21* mutation. In *stag2b*$^{-/-}$ much more complex interactions were observed between the genotype and BIO treatment. In the dendrograms of genes differentially responsive to BIO (539 genes), untreated *stag2b*$^{-/-}$ gene sets cluster with BIO-treated wild type, and BIO-treated *stag2b*$^{-/-}$ gene sets cluster with untreated wild type (Fig. 8C). This suggests that there is an altered baseline of Wnt signaling in *stag2b*$^{-/-}$ and also that BIO stimulation normalizes the expression of select dysregulated genes in *stag2b*$^{-/-}$ tailbuds.

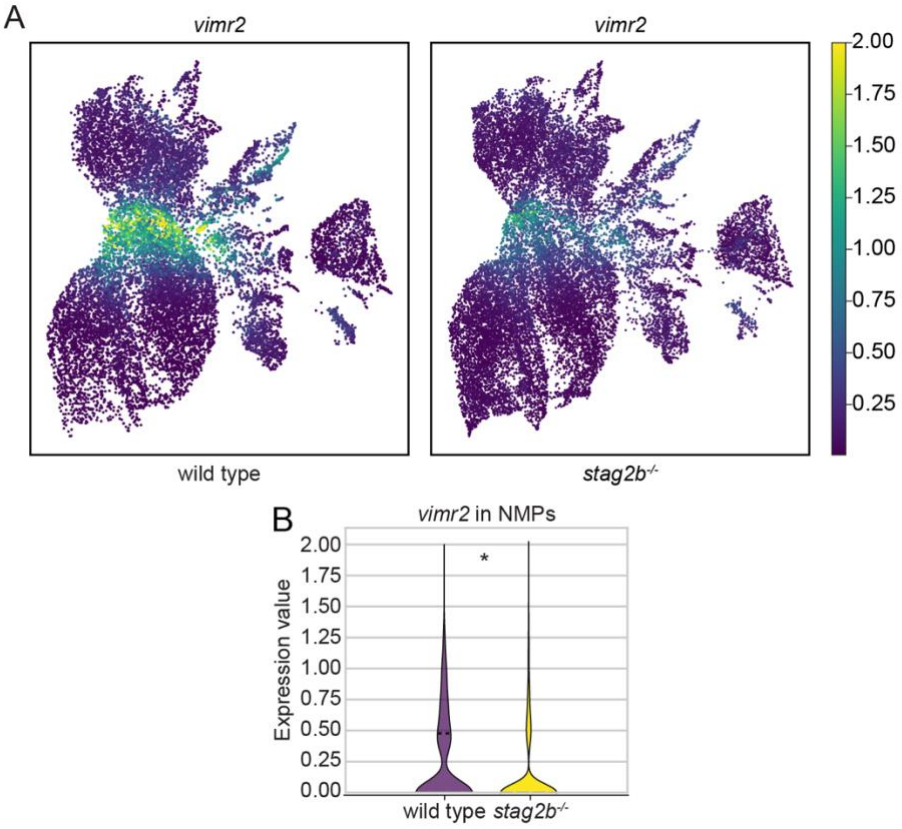

Fig. 9. *stag2b* mutation affects *vimr2* expression in NMPs. (A) Expression of *vimr2* in UMAP representation in wild type and *stag2b*$^{-/-}$ tailbuds at 16 somites. **(B)** Violin plot showing downregulation of *vimr2* expression in the NMPs in *stag2b*$^{-/-}$. * p <0.005, Wilcoxon rank-sum test.

A notable example of a gene with expression that is rescued by BIO in *stag2b*$^{-/-}$ but not in *rad21* mutants is *vimr2*, a marker of epithelial-to-mesenchymal transition (EMT) and mesoderm formation in the tailbud [85]. Expression of *vimr2* was strongly downregulated in all cohesin mutant tailbuds (Fig. 8D, Fig. S9). While BIO treatment had a minimal effect on *vimr2* transcript counts in wild type and *rad21* mutants, it restored *vimr2* levels in *stag2b*$^{-/-}$ tailbuds to wild type (Fig. 8D). Interestingly, our single cell RNA-seq data show that *vimr2* is expressed in NMPs and is significantly downregulated in *stag2b*$^{-/-}$ mutants (Fig. 9AB). This finding raises the possibility that EMT anomalies marked by downregulated *vimr2* could be responsible for changes in mesoderm induction in *stag2b*$^{-/-}$.

Wnt stimulation rescues notochord width in *stag2b*$^{-/-}$ tailbuds

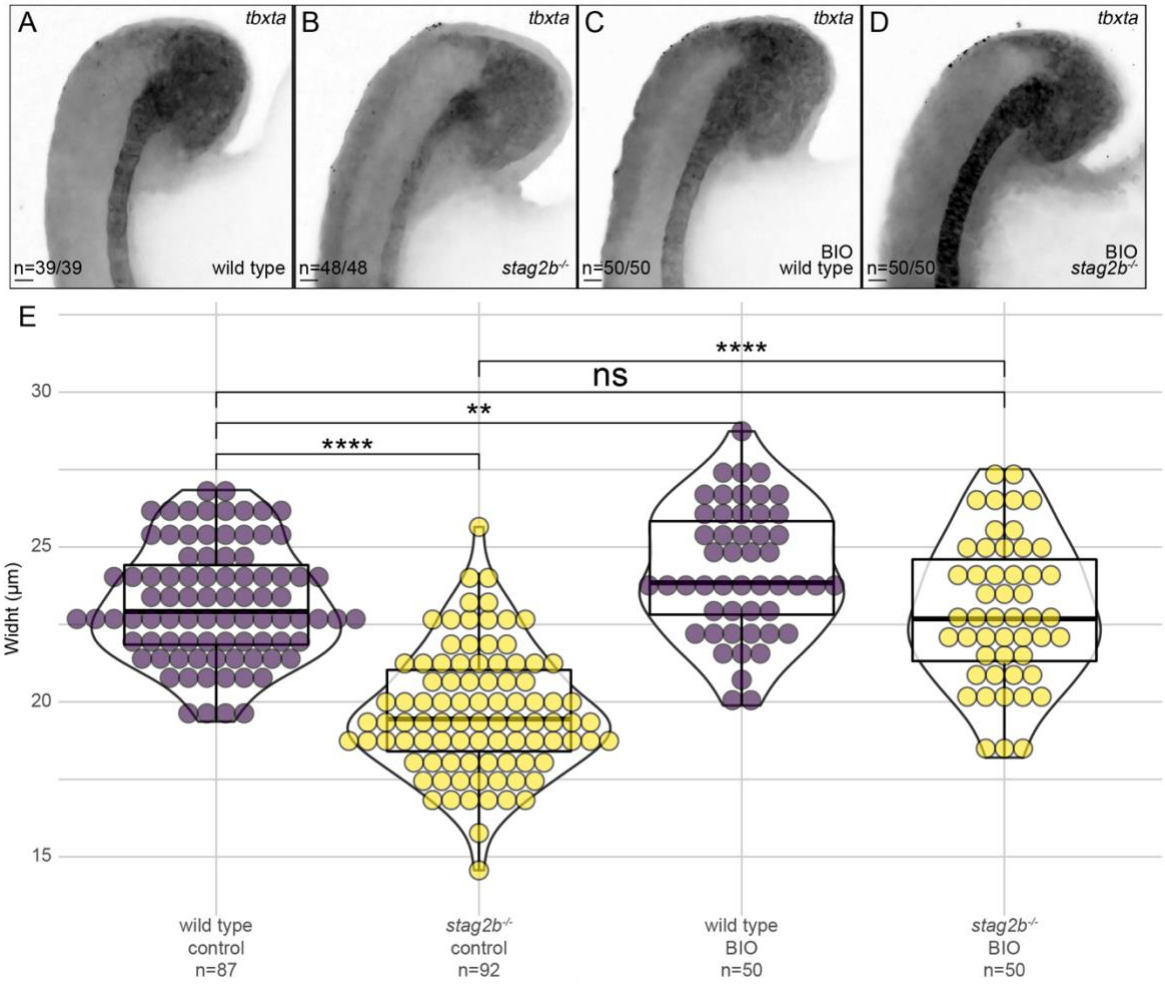

Fig. 10. Wnt stimulation rescues notochord width in *stag2b*$^{-/-}$. (A-D) Expression pattern of *tbxta* in wild type (A, C) and *stag2b*$^{-/-}$ (B, D) zebrafish tailbuds with (C, D) and without (A, B) Wnt stimulation. Images are maximum intensity projections of 3 (4.8 µm) optical sections. Scale bars are 20 µm. The number of embryos with each expression pattern out of the total analyzed is noted. **(E)** Violin plots with overlaid box plots visualizing measurements of notochord width. The x-axis indicates the genotype, treatment status and the number of embryos measured in each group. Significance was determined using an unpaired t-test: ** $p < 0.01$, **** $p < 0.0001$.

If BIO stimulation can restore transcription in *stag2b*$^{-/-}$ tailbuds, we reasoned that it may also rescue the narrower notochord phenotype in *stag2b*$^{-/-}$ embryos. Using HCR with probes for *sox2*, *tbxta*, and *tbx16*, we quantified the thickness of the notochord (*tbxta*) in wild-type and *stag2b*$^{-/-}$ embryos both with and without BIO treatment (Fig. 10, Fig. S10). While Wnt stimulation modestly increased notochord width in wild type ($p \leq 0.01$), it significantly increased the width in *stag2b*$^{-/-}$ embryos ($p \leq 0.0001$) (Fig. 10E). When we compared the notochord width in wild type embryos to that in *stag2b*$^{-/-}$ embryos treated with BIO, the difference was statistically insignificant. Therefore, Wnt stimulation rescues the notochord phenotype in *stag2b*$^{-/-}$ embryos.

In summary, our findings suggest that Wnt stimulation with BIO normalizes gene expression and phenotype in *stag2b*$^{-/-}$ embryos, while gene expression in *rad21* mutants is unresponsive to Wnt stimulation.

Discussion

Germline mutations in subunits of the cohesin complex or its regulators are implicated in developmental disorders known as cohesinopathies, and somatic mutations are now known to cause a variety of cancers. Such cohesin mutations are invariably partial loss-of-function rather than null alleles, because of the essential cell cycle role of cohesin. To date, few studies have compared the developmental consequences of reducing the overall amount of cohesin versus altering its composition. Upon deficiency of Stag2, the Stag1 subunit will compensate in the cohesin complex thereby altering cohesin composition. Upon deficiency of Rad21, the overall quantity of cohesin complexes on DNA decreases. In this study, we took advantage of *stag* and *rad21* gene mutations in zebrafish to show that cohesin composition versus quantity lead to strikingly different consequences for gene transcription and cell differentiation.

Rad21 is an essential subunit in the cohesin complex. Using a zebrafish point mutation *rad21*^{nz171} that progressively reduces *rad21* transcript from heterozygotes to homozygotes, we show that Rad21 deficiency dose-dependently correlates with downregulation of core cohesin subunits, a transcriptional dysregulation signature enriched in cell cycle genes, and a block in mesoderm induction in the tailbud. Although *rad21* heterozygotes are viable and fertile, developmental anomalies in the tailbud are more similar between heterozygotes and *rad21* homozygotes (which die by 48 hpf) than homozygous viable *stag2b*$^{-/-}$. Additional mutation of *stag1b* on top of *stag2b* resulted in loss of viability and a phenotype resembling *rad21* mutants.

It is not clear whether the consequences of Rad21 deficiency are related to cell cycle effects. A recent study showed that blocking the cell cycle in zebrafish does not affect the

development of cell types [86], although it does affect the numbers of presomitic mesoderm cells and erythrocytes [86]. Consistent with cell cycle effects, we observed that *rad21* mutation impacts mesoderm differentiation and our previous work has shown that erythropoiesis is downregulated in cohesin mutants [12, 77]. However, cell cycle impairment is unlikely to account for all the defects associated with the reduction in cohesin dose, and it does not explain the transcriptional and phenotypic changes observed in a viable, fertile *rad21* heterozygotes. Our previous work has shown that Rad21 deficiency has remarkably specific transcriptional and developmental consequences, for example, cell-type specific loss of *runx1* expression [12, 78, 87]. Consistent with a non-cell cycle related transcriptional role for Rad21, complete removal of Rad21 interferes with transcription in post-mitotic neurons, which is rescued upon restoring Rad21 [88].

In contrast, homozygous mutation in *stag2b* has no discernible effect on the cell cycle. Although *stag2b*$^{-/-}$ mutants are viable and fertile, changes in mesoderm differentiation are apparent at tailbud stages. Single cell sequencing of tailbuds shows that the proportions of mesoderm in the tailbud are subtly altered, and the EMT driver *vimr2* is downregulated in NMPs. Genes in the Wnt signaling pathway are altered in NMPs and in paraxial mesoderm. The noticeably narrower notochord in *stag2b*$^{-/-}$ mutants is rescued by stimulation of Wnt signaling via inhibition of GSK3. Moreover, transcription in *stag2b*$^{-/-}$ mutant tailbuds is rescued to wild type upon GSK3 inhibition. Our results suggest that disrupted Wnt signaling likely accounts for transcriptional and phenotypic changes in the *stag2b*$^{-/-}$ mutants. Notably, Wnt stimulation does not rescue transcription in *rad21* genotypes. This suggests that reduced cohesin dose is epistatic to altered cohesin composition.

We and others have previously reported dysregulated Wnt signaling upon cohesin mutation [78, 89-93] but the directionality of Wnt signaling disturbance remains unclear. We have shown stabilization of β-catenin and both up- and downregulation of components of the Wnt signaling pathway, indicating that the effects of cohesin deficiency on Wnt are likely to be complex [90]. Interestingly, GSK3α inhibition was shown to stabilize cohesin on chromatin, promoting continued loop extrusion [94]. Stabilized loop extrusion is dependent on cohesin as it was eliminated with knockdown of Rad21 [94]. The compound (BIO) we used to stimulate Wnt inhibits GSK3 and does not distinguish between α and β forms. It is possible that GSK3 inhibition was able to rescue transcription and phenotypes in *stag2b*$^{-/-}$ mutant tailbuds but not in *rad21* mutants because a reduction in Rad21 reduces the number of complexes that can be stabilized, whereas Stag1b compensates for loss of Stag2b in those complexes.

Stag1-containing cohesin resides primarily at CTCF sites that demarcate contact domains that are invariant between tissues. Stag2-containing cohesin resides at CTCF and non-CTCF sites, where it is thought to regulate tissue-specific transcription [16, 65, 69, 83]. Viny et al (2019) showed that Stag1-cohesin cannot fully substitute for Stag2 cohesin in haematopoietic stem cells. It is possible that Stag1-containing cohesin has different properties in loop extrusion than Stag2 cohesin [16, 95] and likely that developmental gene transcription in *stag2b*$^{-/-}$ mutant tailbuds is altered because of compensation by Stag1b. We do not believe the other Stag orthologs are major contributors to development in zebrafish owing to the lethality of *stag1b*$^{-/-}$; *stag2b*$^{-/-}$ double mutants.

Stag proteins may have functions that are independent of the cohesin complex. For example a recent study found that upon RAD21 depletion, STAG proteins remain bound to chromatin, interact with CTCF, and cluster in 3D [67]. STAG proteins interact with RNA and R-loops even in the absence of cohesin. Drosophila's SA cohesin subunit (equivalent to Stag2) is differentially enriched at enhancers and promoters near origins of replication where it is proposed to recruit cohesin [70]. In contrast, RAD21 appears to be key for stable binding of cohesin at CTCF sites. A recent study showed that when RAD21 is cleaved, cohesin is released from DNA including at CTCF sites and loops at these elements are lost [55]. Interestingly CTCF-independent cohesin-anchored loops within chromatin domains persisted despite RAD21 cleavage. The different molecular and structural behaviour of RAD21 and STAG proteins is consistent with the diverse developmental consequences we observed upon germline mutation in these genes.

It is possible that some of the molecular basis for developmental abnormalities is shared between NIPBL deficiency and STAG2 mutation. A recent preprint describing single cell RNA sequencing of early-stage mouse embryos with one deleted copy of *Nipbl* showed that these embryos also experience changes in mesoderm fate and have altered mesoderm cell populations [96]. Nipbl loss altered the regulation of genes involved in EMT, which parallels our findings in *stag2b* mutant zebrafish embryos. This raises the possibility that abnormal regulation of mesoderm fate could be a conserved feature of the cohesinopathies.

Alternate transcriptional and developmental consequences with Stag2 and Rad21 deficiency have implications for the amelioration of cohesinopathies where Wnt agonists have been explored as potential therapeutic agents for individuals with CdLS [91], and additionally, for the treatment of cohesin-mutant cancers [90]. Our results suggest that reducing cohesin dose has very different consequences to altering cohesin composition. This indicates that mutations in core

cohesin subunits need to be considered differently to mutations in alternate cohesin subunits or cohesin regulators when developing therapeutics.

Materials and Methods

Zebrafish Husbandry

Wild type (WIK) [97], *stag1a*^{nz204} [77], *stag1b*^{nz205} [77], *stag2b*^{nz207} [77] and *rad21*^{nz171} *[12]* zebrafish lines were maintained at 28 °C according to established husbandry methods [98]. Zebrafish were housed in the Otago Zebrafish Facility (Department of Pathology, University of Otago, Dunedin, New Zealand). All animal work was performed in accordance with the Otago Zebrafish Facility Standard Operating Procedures (AUP 21-110) and under Environmental Risk Management Authority approval numbers GMC005627, GMD100922 and GMC001366. For all experiments, embryos were developed at 22 or 28 °C.

Whole-mount *in situ* hybridisations (WISH) and hybridization chain reaction (HCR) RNA-FISH

WISH for *runx1* was performed using 0.5 ng/µL of riboprobe as previously described [82]. Probes for *sox2*, *tbxta*, and *tbx16* and HCR reagents were purchased from Molecular Instruments, Inc (USA). HCR was performed according to the manufacturer's protocol for zebrafish embryos.

Flow cytometry

Embryos at the 16-somite stage were fixed in methanol [99] and tailbuds were dissected (n=30). *rad21* heterozygotes and homozygotes were identified by genotyping the heads of individual embryos using genomic DNA extraction [100] followed by a custom TaqMan assay. Cells were filtered through a 40 µm cell strainer and nuclei were stained with DRAQ5 (#ab108410, Abcam) at 5 µM final concentration on ice for 45 min in the dark. Cell cycle profiles of three independent replicates for each genotype were obtained using a BD FACS Aria III (BD Biosciences). Data analysis and plots were generated using Cytoflow [101].

BrdU incorporation

For cell cycle analyses dechorionated embryos were incubated in 10 mM BrdU in Ringer's solution for 30 minutes on ice, rinsed three times with Ringer's solution and incubated for 30 minutes or 2 hours at 28 °C. Embryos were fixed with 4% PFA overnight at 4 °C, dehydrated in methanol and stored at -20 °C in 100% MeOH.

For staining, embryos were rehydrated in a series of 5-minute washes with PBST/MeOH. Embryos older than 24 hpf were treated with 10 µg/ml proteinase K for 10 minutes, followed by three 5-minute washes in PBST and post-fixation in 4% PFA for 20 minutes at room temperature. Samples were then rinsed three times with sterile distilled water. For BrdU staining, embryos were rinsed twice in 2 N HCl and incubated in 2 N HCl for 1 hour at room temperature to denature DNA and expose the BrdU epitope. Alternatively, for antigen exposure, embryos were treated with acetone for 20 minutes on ice.

Immunohistochemistry

Samples were rinsed twice with sterile distilled water and washed twice with PBST for 5 minutes. Embryos were incubated in blocking solution (0.2% Roche block, 10% FBS, 1% DMSO in PBST) for 30 minutes, followed by a 2-day incubation with primary antibodies at 4 °C. Primary antibodies used are as follows: anti-phH3 (#3377, Cell Signaling Technology; 1:1000), anti-α-tubulin (#T6199, Sigma-Aldrich; 1:500), and anti-BrdU (#B35141, Thermo Fisher Scientific). Antibodies were washed off with three 10-minute washes in PBST and two 10-minute washes in 1% FBS in PBST. Embryos were then incubated with secondary antibodies (1:1000) in 1% FBS in PBST at 4 °C for 2 days in the dark. Secondary antibodies used for immunofluorescence were goat anti-mouse Alexa Fluor 488 (1:1000, #A11001, Thermo Fisher Scientific), chicken anti-rabbit Alexa Fluor 647(1:1000, #A21443, Thermo Fisher Scientific). On the second day, Hoechst 33342 (1 µg/ml) (Thermo Fisher Scientific; 1:1000) was added. Embryos were washed five times for 10 minutes with PBST and imaged.

Microscopy

Fixed embryos were immersed in 70% glycerol to obtain bright field images. Live embryos were anaesthetized with MS-222 (200 mg/L) and embedded in 3% methylcellulose. Bright-field images were captured using the Leica M205FA epifluorescence microscope equipped with a DFC490 camera and Leica Applications Suite software (Leica Microsystems, Germany).

For confocal microscopy embryos were mounted in 1% low melting agarose (w/v). Confocal images were acquired using a Nikon C2 confocal microscope as Z-stacks of the optical

sections. The images were processed using NIS-Elements Denoise.ai Software. Maximum intensity projections were used for the figures.

BIO treatment

30 μM 6-bromoindirubin-3'-oxime (BIO) solution was diluted to 2.5 μM BIO in E3 medium. Embryos were sorted into 50 embryos per plate and treated with 2.5 μM BIO from 4 hpf until tailbud dissection at the 16-somite stage.

Tailbud bulk RNA sequencing (RNA-seq) and analyses

Tailbuds were dissected from stage-matched embryos at 16 somites (16-18 hpf) as illustrated in Figure 3A. For RNA-seq, tailbuds were individually lysed in 3 μL of RLT + BME (Qiagen RNeasy) and stored in separate PCR tubes at -80 °C to await genotyping of heads (for *rad21*$^{-/-}$ and *rad21*$^{+/-}$). Total RNA was extracted from the pools of 80 tailbuds per sample using the RNeasy Micro kit (74104; Qiagen, Germany). Quality and quantity of RNA were assessed using Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA), Agilent RNA 6000 Nano Kit on 2100 Bioanalyzer (Agilent Technologies, Netherlands) and NanoPhotometer NP80 Touch (Implen GmbH, Germany).

Libraries were prepared from 250 ng of total RNA using the TruSeq Stranded mRNA Library Prep kit (Illumina, USA) and TruSeq RNA CD Index Plate (Illumina, USA) for sample multiplexing. The concentration of the libraries was quantified using a Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA), and the mean fragment size was assessed using the DNA High Sensitivity KIT on a 2100 Bioanalyzer (Agilent Technologies, Netherlands). 4 nM pooled libraries was sequenced on NovaSeq S1 flow cell by Livestock Improvement Corporation Ltd. (New Zealand).

RNA-seq reads were trimmed using Cutadapt [102], and aligned to the reference genome (GRCz11) with HISAT2 [103] and SAMtools [104]. FeatureCounts [105] was used to generate fragment count matrices. DESeq2 [106] was used to perform differential gene expression analysis, and multi-testing correction was done using the Benjamini-Hochberg procedure. The false discovery rate (FDR) threshold was set at 5%. Pathway analysis was performed using Reactome [107] and Metascape [108]. Genotype-specific BIO treatment effects were tested by adding an interaction term (modelling the interaction between treatment and genotype) at the experimental design stage prior to calling differential genes.

Single cell RNA sequencing

Stage matched 16-somite embryos (wild type or *stag2b*$^{-/-}$) were dechorionated using pronase (20 mg/ml in E3) and deyolked in calcium-free Ringer's solution (116 mM NaCl, 2.6 mM KCl, 5 mM HEPES, pH 7.0). Tailbud tissue was dissected using a small needle and pooled (n=30). Tailbuds were then incubated with collagenase (2 mg/ml in 0.05% trypsin, 1 mM EDTA, pH 8.0, PBS) in 1.3 ml volume at 28 °C for 15 mins with intermittent pipetting to achieve a single cell suspension. The reaction was stopped by adding 200 μL of a stop solution (30% calf serum, 6 mM $CaCl_2$, PBS). Cells were centrifuged at 500 x *g* for 5 min and re-suspended in 1 ml of resuspension solution (1% Calf Serum, 0.8 mM $CaCl_2$, 50 U/ml Penicillin, 0.05 mg/ml Streptomycin). After centrifuging cells were resuspended in 700 μL of resuspension buffer and filtered through a 40 μm cell strainer and kept on ice. Single cell suspensions were processed at the Genomics High Throughput Facility (UC Irvine, USA) according to the manufacturers protocol for the 10X Chromium single cell platform (10X Genomics, Pleasanton, CA. USA) specifically the Chromium Single Cell 3' Library and Gel Bead Kit v3 (Cat.No. PN-1000128). Libraries were sequenced on a HiSeq2500 platform (Illumina) yielding 1,024,641,718 reads for wild type and 1,166,072,985 reads for the *stag2b*$^{-/-}$ sample.

Single cell RNA sequencing data analysis

Single-cell RNA-seq FASTQ files were demultiplexed using Cellranger (v7.1.0) [109], mapped to the Danio rerio.GRCz11 (danRer11) transcriptome (v4.3.2) [110] including intronic reads. We obtained an estimated number of 18,704 cells (wild type) and 26,360 cells (*stag2b*$^{-/-}$). In the wild type sample mean reads per cell were 54,782, median UMI (unique molecular identifier) were 15,696 and 3,651 genes detected per cell.

In *stag2b*$^{-/-}$ we detected 44,236 mean reads per cell, 13,922 median UMIs and 3,561 median genes per cell. The total number of genes detected was 25,879 for wild type and 26,168 for *stag2b*$^{-/-}$. 98.2% and 98.1% of reads (wild type/ *stag2b*$^{-/-}$) had valid barcodes with Q30 of 94.4% for both samples. 92% and 91.5% of the reads mapped confidently to the zebrafish genome. Downstream anlaysis was performed using scvi-tools (v0.20.3) [111] and scanpy (v1.9.3) [112]. After filtering of empty cells doublet removal was performed using Solo [113]. scanpy was used to filter out cells with less than 200 genes, genes detected in less than 3 cells, cells exceeding gene counts of 98% of the median. We also filtered out cells having more than 15% mitochondrial reads (indicating cellular stress).

Filtering and QC steps resulted in 15,298 wild type cells and 21,278 *stag2b*$^{-/-}$cells. The data were then normalised to 10,000 UMIs per cell using the function (scanpy.pp.normalize). Integration of the two datasets was performed using scvi-tools function scVI model [114]. Specifically we set up the model using the following command: scvi.model.SCVI.setup_anndata (anndata, layer = "counts", categorical_covariate_keys=["Genotype"], continuous_covariate_keys=['pct_counts_mt', 'total_counts']). We then trained the model and obtained the latent space with model.get_latent_representation(). The neighbourhood graph as well as the UMAP (Uniform Manifold Approximation and Projection) plot was determined using scanpy functions using the scVI latent space as input. Leiden clustering (resolution 0.5) resulted in 16 clusters (data not shown), which were partially merged into common clusters based on the expression of canonical marker genes. Significant differentially expressed genes (DEGs) between clusters and genotypes were determined using the the model calculated above using scVI as well as the Wilcoxon rank-sum test (Benjamini-Hochberg correction).

A list of differentially expressed genes in NMPs of wild type and *stag2b*$^{-/-}$ tailbuds can be found in Data S4. We used Augur (Skinnider et al., 2021) to analyse the effect of on cell types based on the genotype. Embeddings were calculated using scanpy function (scanpy.tl.embedding_density). We used the function (scanpy.tl.score_genes) for a list of genes obtained from wiki pathways: Canonical_Wnt_pathway (WP1349). The function (scanpy.pp.scale) was used to scale the data to unit variance and zero mean and the Wilcoxon rank-sum test (Benjamini-Hochberg correction) was used to determine significance for the data shown in Fig. 7.

Data availability

Bulk RNA-seq data are available at GEO under accession number GSE247246.

Raw and processed sc-RNA-seq data generated in this study are available at GEO under GSE171482.

References

1. Nasmyth, K. and C.H. Haering, *Cohesin: its roles and mechanisms.* Annu Rev Genet, 2009. **43**: p. 525-58.
2. Huis, P.J., et al., *Characterization of a DNA exit gate in the human cohesin ring.* Science, 2014. **346**(6212): p. 968-972.
3. Gruber, S., C.H. Haering, and K. Nasmyth, *Chromosomal cohesin forms a ring.* Cell, 2003. **112**(6): p. 765-77.

4. Horsfield, J.A., C.G. Print, and M. Monnich, *Diverse developmental disorders from the one ring: distinct molecular pathways underlie the cohesinopathies.* Front Genet, 2012. **3**: p. 171.

5. Dorsett, D. and L. Strom, *The ancient and evolving roles of cohesin in gene expression and DNA repair.* Curr Biol, 2012. **22**(7): p. R240-50.

6. Ciosk, R., et al., *Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins.* Molecular cell, 2000. **5**(2): p. 243-254.

7. Kueng, S., et al., *Wapl controls the dynamic association of cohesin with chromatin.* Cell, 2006. **127**(5): p. 955-967.

8. Sjögren, C. and K. Nasmyth, *Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae.* Current Biology, 2001. **11**(12): p. 991-995.

9. Watrin, E. and J.-M. Peters, *Cohesin and DNA damage repair.* Experimental cell research, 2006. **312**(14): p. 2687-2693.

10. Hou, W., et al., *Cohesin in DNA damage response and double-strand break repair.* Critical Reviews in Biochemistry and Molecular Biology, 2022. **57**(3): p. 333-350.

11. Cukrov, D., et al., *Antioxidant treatment ameliorates phenotypic features of SMC1A-mutated Cornelia de Lange syndrome in vitro and in vivo.* Human molecular genetics, 2018. **27**(17): p. 3002-3011.

12. Horsfield, J.A., et al., *Cohesin-dependent regulation of Runx genes.* Development, 2007. **134**(14): p. 2639-49.

13. Percival, S.M., et al., *Variations in dysfunction of sister chromatid cohesion in esco2 mutant zebrafish reflect the phenotypic diversity of Roberts syndrome.* Dis Model Mech, 2015. **8**(8): p. 941-55.

14. Gerlich, D., et al., *Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication.* Current biology, 2006. **16**(15): p. 1571-1578.

15. Horsfield, J.A., *Full circle: a brief history of cohesin and the regulation of gene expression.* The FEBS journal, 2023. **290**(7): p. 1670-1687.

16. Cuadrado, A. and A. Losada, *Specialized functions of cohesins STAG1 and STAG2 in 3D genome architecture.* Curr Opin Genet Dev, 2020. **61**: p. 9-16.

17. Nishiyama, T., *Cohesion and cohesin-dependent chromatin organization.* Current opinion in cell biology, 2019. **58**: p. 8-14.

18. Ball, A.R., Jr., Y.Y. Chen, and K. Yokomori, *Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies.* Biochim Biophys Acta, 2014. **1839**(3): p. 191-202.

19. Zhu, Z. and X. Wang. *Roles of cohesin in chromosome architecture and gene expression*. in *Seminars in Cell & developmental biology*. 2019. Elsevier.

20. Sanborn, A.L., et al., *Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes.* Proceedings of the National Academy of Sciences, 2015. **112**(47): p. E6456-E6465.

21. Fudenberg, G., et al., *Formation of chromosomal domains by loop extrusion.* Cell reports, 2016. **15**(9): p. 2038-2049.

22. Davidson, I.F., et al., *DNA loop extrusion by human cohesin.* Science, 2019. **366**(6471): p. 1338-1345.

23. Krijger, P.H. and W. de Laat, *Regulation of disease-associated gene expression in the 3D genome.* Nat Rev Mol Cell Biol, 2016. **17**(12): p. 771-782.

24. Hnisz, D., et al., *Activation of proto-oncogenes by disruption of chromosome neighborhoods.* Science, 2016. **351**(6280): p. 1454-1458.

25. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell, 2014. **159**(7): p. 1665-1680.

26. Mayerova, N., L. Cipak, and J. Gregan, *Cohesin Biology: From Passive Rings to Molecular Motors.* Trends Genet, 2020. **36**(6): p. 387-389.

27. Wendt, K.S., et al., *Cohesin mediates transcriptional insulation by CCCTC-binding factor.* Nature, 2008. **451**(7180): p. 796-801.

28. Parelho, V., et al., *Cohesins functionally associate with CTCF on mammalian chromosome arms.* Cell, 2008. **132**(3): p. 422-433.

29. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-380.

30. Meier, M., et al., *Cohesin facilitates zygotic genome activation in zebrafish.* Development, 2018. **145**(1): p. dev156521.

31. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture.* Nature, 2010. **467**(7314): p. 430-435.

32. Ochi, Y., et al., *Combined Cohesin-RUNX1 Deficiency Synergistically Perturbs Chromatin Looping and Causes Myelodysplastic Syndromes.* Cancer Discov, 2020. **10**(6): p. 836-853.

33. Marsman, J. and J.A. Horsfield, *Long distance relationships: enhancer-promoter communication and dynamic gene transcription.* Biochim Biophys Acta, 2012. **1819**(11-12): p. 1217-27.

34. Friman, E.T., et al., *Ultra-long-range interactions between active regulatory elements.* Genome Research, 2023. **33**(8): p. 1269-1283.

35. Goel, V.Y., M.K. Huseyin, and A.S. Hansen, *Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments.* Nature Genetics, 2023: p. 1-9.

36. Kane, L., et al., *Cohesin is required for long-range enhancer action at the Shh locus.* Nature structural & molecular biology, 2022. **29**(9): p. 891-897.

37. Piché, J., et al., *The expanding phenotypes of cohesinopathies: one ring to rule them all!* Cell Cycle, 2019. **18**(21): p. 2828-2848.

38. Opitz, J.M. and J.F. Reynolds, *The Brachmann-de Lange syndrome.* American journal of medical genetics, 1985. **22**(1): p. 89-102.

39. Ireland, M., D. Donnai, and J. Burn, *Brachmann-de Lange syndrome. Delineation of the clinical phenotype.* American journal of medical genetics, 1993. **47**(7): p. 959-964.

40. Jackson, L., et al., *de Lange syndrome: a clinical review of 310 individuals.* American journal of medical genetics, 1993. **47**(7): p. 940-946.

41. Liu, J., et al., *Transcriptional dysregulation in NIPBL and cohesin mutant human cells.* PLoS biology, 2009. **7**(5): p. e1000119.

42. Krantz, I.D., et al., *Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B.* Nature genetics, 2004. **36**(6): p. 631-635.

43. Tonkin, E.T., et al., *NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome.* Nature genetics, 2004. **36**(6): p. 636-641.

44. Cheung, K. and J. Upton, *Cornelia de Lange Syndrome.* J Hand Surg Am, 2015. **40**(12): p. 2501-3.

45. Deardorff, M.A., S.E. Noon, and I.D. Krantz, *Cornelia de Lange syndrome.* 2020.

46. Deardorff, M.A., et al., *HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle.* Nature, 2012. **489**(7415): p. 313-7.

47. Deardorff, M.A., et al., *Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation.* Am J Hum Genet, 2007. **80**(3): p. 485-94.

48. Deardorff, M.A., et al., *RAD21 mutations cause a human cohesinopathy.* The American Journal of Human Genetics, 2012. **90**(6): p. 1014-1027.

49. Mungan, Z., et al., *Familial visceral myopathy with pseudo-obstruction, megaduodenum, Barrett's esophagus, and cardiac abnormalities.* Official journal of the American College of Gastroenterology| ACG, 2003. **98**(11): p. 2556-2560.

50. Zhang, B.N., et al., *A cohesin subunit variant identified from a peripheral sclerocornea pedigree.* Disease Markers, 2019. **2019**.

51. Kruszka, P., et al., *Cohesin complex-associated holoprosencephaly.* Brain, 2019. **142**(9): p. 2631-2643.

52. Krab, L.C., et al., *Delineation of phenotypes and genotypes related to cohesin structural protein RAD21.* Human genetics, 2020. **139**(5): p. 575-592.

53. Cheng, H., N. Zhang, and D. Pati, *Cohesin subunit RAD21: From biology to disease.* Gene, 2020. **758**: p. 144966.

54. Sun, Y., et al., *RAD21 is the core subunit of the cohesin complex involved in directing genome organization.* Genome Biology, 2023. **24**(1): p. 1-27.

55. Liu, Y. and J. Dekker, *CTCF–CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity.* Nature cell biology, 2022. **24**(10): p. 1516-1527.

56. Nagasaka, K., et al., *Cohesin mediates DNA loop extrusion and sister chromatid cohesion by distinct mechanisms.* Molecular Cell, 2023. **83**(17): p. 3049-3063. e6.

57. Cratsenberg, D.M., P.J. Winningham, and L.J. Starr, *Second reported individual with a partial STAG2 deletion: Middle interhemispheric variant holoprosencephaly in STAG2-related cohesinopathy.* Clinical Dysmorphology, 2021. **30**(3): p. 159-163.

58. Soardi, F.C., et al., *Familial STAG2 germline mutation defines a new human cohesinopathy.* NPJ genomic medicine, 2017. **2**(1): p. 7.

59. Mullegama, S.V., et al., *Mutations in STAG2 cause an X-linked cohesinopathy associated with undergrowth, developmental delay, and dysmorphia: Expanding the phenotype in males.* Molecular genetics & genomic medicine, 2019. **7**(2): p. e00501.

60. Mullegama, S.V., et al., *De novo loss-of-function variants in STAG2 are associated with developmental delay, microcephaly, and congenital anomalies.* American journal of medical genetics Part A, 2017. **173**(5): p. 1319-1327.

61. Freyberger, F., et al., *Expanding the known phenotype of Mullegama–Klein–Martinez syndrome in male patients.* Human Genome Variation, 2021. **8**(1): p. 37.

62. Yuan, B., et al., *Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies.* Genetics in Medicine, 2019. **21**(3): p. 663-675.

63. Schmidt, J., et al., *Somatic mosaicism in STAG2-associated cohesinopathies: Expansion of the genotypic and phenotypic spectrum.* Frontiers in Cell and Developmental Biology, 2022. **10**: p. 1025332.

64. Gokce-Samar, Z., et al., *STAG2 microduplication in a patient with eyelid myoclonia and absences and a review of EMA-related reported genes.* European Journal of Medical Genetics, 2022: p. 104636.

65. Kojic, A., et al., *Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization.* Nature structural & molecular biology, 2018. **25**(6): p. 496-504.

66. Cheng, N., et al., *STAG2 promotes the myelination transcriptional program in oligodendrocytes.* Elife, 2022. **11**: p. e77848.

67. Porter, H., et al., *Cohesin-independent STAG proteins interact with RNA and R-loops and promote complex loading.* Elife, 2023. **12**: p. e79386.

68. Remeseiro, S., et al., *A unique role of cohesin-SA1 in gene regulation and development.* EMBO J, 2012. **31**(9): p. 2090-102.

69. Viny, A.D., et al., *Cohesin Members Stag1 and Stag2 Display Distinct Roles in Chromatin Accessibility and Topological Control of HSC Self-Renewal and Differentiation.* Cell Stem Cell, 2019. **25**(5): p. 682-696 e8.

70. Pherson, M., et al., *Cohesin occupancy and composition at enhancers and promoters are linked to DNA replication origin proximity in Drosophila.* Genome research, 2019. **29**(4): p. 602-612.

71. Bailey, M.L., et al., *Paralogous synthetic lethality underlies genetic dependencies of the cancer-mutated gene STAG2.* Life Science Alliance, 2021. **4**(11).

72. Adane, B., et al., *STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma.* Cancer cell, 2021. **39**(6): p. 827-844. e10.

73. Liu, J. and I.D. Krantz, *Cohesin and human disease.* Annu. Rev. Genomics Hum. Genet., 2008. **9**: p. 303-320.

74. McNairn, A.J. and J.L. Gerton, *The chromosome glue gets a little stickier.* Trends in genetics, 2008. **24**(8): p. 382-389.

75. Dorsett, D. and I.D. Krantz, *On the molecular etiology of Cornelia de Lange syndrome.* Annals of the New York Academy of Sciences, 2009. **1151**(1): p. 22-37.

76. Muto, A., et al., *Multifactorial origins of heart and gut defects in nipbl-deficient zebrafish, a model of Cornelia de Lange Syndrome.* PLoS biology, 2011. **9**(10): p. e1001181.

77. Ketharnathan, S., A. Labudina, and J.A. Horsfield, *Cohesin Components Stag1 and Stag2 Differentially Influence Haematopoietic Mesoderm Development in Zebrafish Embryos.* Front Cell Dev Biol, 2020. **8**: p. 617545.

78. Schuster, K., et al., *A neural crest origin for cohesinopathy heart defects.* Hum Mol Genet, 2015. **24**(24): p. 7005-16.

79. Muto, A. and T.F. Schilling, *Zebrafish as a model to study cohesin and cohesinopathies.* Cohesin and Condensin: Methods and Protocols, 2017: p. 177-196.

80. Row, R.H., et al., *The zebrafish tailbud contains two independent populations of midline progenitor cells that maintain long-term germ layer plasticity and differentiate in response to local signaling cues.* Development, 2016. **143**(2): p. 244-254.

81. Steventon, B. and B. Martin, *A fishy tail: Insights into the cell and molecular biology of neuromesodermal cells from zebrafish embryos.* 2022.

82. Kalev-Zylinska, M.L., et al., *Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis.* 2002.

83. Casa, V., et al., *Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcriptional control.* Genome Research, 2020. **30**(4): p. 515-527.

84. Thisse, B., et al., *Expression of the zebrafish genome during embryogenesis. 2001.* ZFIN Direct Data Submission (http://zfin. org), 2001.

85. Goto, H., et al., *FGF and canonical Wnt signaling cooperate to induce paraxial mesoderm from tailbud neuromesodermal progenitors through regulation of a two-step epithelial to mesenchymal transition.* Development, 2017. **144**(8): p. 1412-1424.

86. Kukreja, K., et al., *Global decoupling of cell differentiation from cell division in early embryo development.* bioRxiv, 2023: p. 2023.07. 29.551123.

87. Marsman, J., et al., *Cohesin and CTCF differentially regulate spatiotemporal runx1 expression during zebrafish development.* Biochim Biophys Acta, 2014. **1839**(1): p. 50-61.

88. Weiss, F.D., et al., *Neuronal genes deregulated in Cornelia de Lange Syndrome respond to removal and re-expression of cohesin.* Nature Communications, 2021. **12**(1): p. 2919.

89. Medina, M.A., et al., *Alternative RUNX1 Promoter Regulation by Wnt/beta-Catenin Signaling in Leukemia Cells and Human Hematopoietic Progenitors.* J Cell Physiol, 2016. **231**(7): p. 1460-7.

90. Chin, C.V., et al., *Cohesin mutations are synthetic lethal with stimulation of WNT signaling.* Elife, 2020. **9**: p. e61405.

91. Grazioli, P., et al., *Lithium as a possible therapeutic strategy for Cornelia de Lange syndrome.* Cell Death Discovery, 2021. **7**(1): p. 34.

92. Pileggi, S., et al., *Cohesin Mutations Induce Chromatin Conformation Perturbation of the H19/IGF2 Imprinted Region and Gene Expression Dysregulation in Cornelia de Lange Syndrome Cell Lines.* Biomolecules, 2021. **11**(11): p. 1622.

93. Mazzola, M., et al., *NIPBL: a new player in myeloid cell differentiation.* Haematologica, 2019. **104**(7): p. 1332.

94. Park, D.S., et al., *High-throughput Oligopaint screen identifies druggable 3D genome regulators.* Nature, 2023: p. 1-9.

95. Alonso-Gil, D., et al., *Different NIPBL requirements of cohesin-STAG1 and cohesin-STAG2.* Nature Communications, 2023. **14**(1): p. 1326.

96. Chea, S., et al., *Gastrulation-stage gene expression in Nipbl+/-mouse embryos foreshadows the development of syndromic birth defects.* bioRxiv, 2023: p. 2023.10. 16.558465.

97. Rauch, G.-J., M. Granato, and P. Haffter, *A polymorphic zebrafish line for genetic mapping using SSLPs on high-percentage agarose gels.* Technical Tips Online, 1997. **2**(1): p. 148-150.

98. Westerfield, M., *The zebrafish book: a guide for the laboratory use of zebrafish (Brachydanio rerio)*. 1995: University of Oregon press.

99. García-Castro, H., et al., *ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics.* Genome biology, 2021. **22**: p. 1-34.

100. Meeker, N.D., et al., *Method for isolation of PCR-ready genomic DNA from zebrafish tissues.* Biotechniques, 2007. **43**(5): p. 610-614.

101. Teague, B., *Cytoflow: A python toolbox for flow cytometry.* bioRxiv, 2022: p. 2022.07.22.501078.

102. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet. journal, 2011. **17**(1): p. 10-12.

103. Kim, D., et al., *Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.* Nature biotechnology, 2019. **37**(8): p. 907-915.

104. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2): p. giab008.

105. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-930.

106. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 1-21.

107. Yu, G. and Q.-Y. He, *ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization.* Molecular BioSystems, 2016. **12**(2): p. 477-479.

108. Zhou, Y., et al., *Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.* Nature communications, 2019. **10**(1): p. 1523.

109. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nature communications, 2017. **8**(1): p. 14049.

110. Lawson, N.D., et al., *An improved zebrafish transcriptome annotation for sensitive and comprehensive detection of cell type-specific genes.* Elife, 2020. **9**: p. e55792.

111. Gayoso, A., et al., *A Python library for probabilistic analysis of single-cell omics data.* Nature biotechnology, 2022. **40**(2): p. 163-166.

112. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome biology, 2018. **19**: p. 1-5.

113. Bernstein, N.J., et al., *Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning.* Cell systems, 2020. **11**(1): p. 95-101. e5.

114. Lopez, R., et al., *Deep generative modeling for single-cell transcriptomics.* Nature methods, 2018. **15**(12): p. 1053-1058.

Acknowledgments

The authors would like to thank Noel Jhinku and Dr Doug Mackie for the expert management of the zebrafish facility and Dr Robert Woolley for the confocal microscopy advice. The authors are grateful to Dr Ben Martin and Dr Christian Mosimann for helpful advice and discussions.

Funding:

Royal Society of NZ Marsden Fund grant 20-UOO-071 (JAH, GG)

Author contributions:

Conceptualization: AAL, JA, JAH
Methodology: AAL, MM, GG, DT, BM, JA, JAH
Investigation: AAL, MM, GG, DT, SK, JA, JAH
Visualization: AAL, MM, GG, DT, SK, TFS, JA, JAH
Supervision: GG, JA, TFS, JAH
Writing—original draft: AAL, MM, JAH
Writing—review & editing: AAL, MM, GG, DT, SK, TFS, JA, JAH

Competing interests:

No competing interest declared

Data and materials availability:

All data are available in the main text or the supplementary materials.

Supplementary Materials

Fig. S1. Trunk haemorrhaging at 48 hpf and *runx1* expression at 12 hpf in *stag1a*$^{-/-}$; *stag2b*$^{-/-}$ embryos.

(A) Lateral views of representative and *stag1a*$^{-/-}$; *stag2b*$^{-/-}$ embryos at 48 hpf. The boxed region (inset) outlines a trunk haemorrhage representative of those observed in around 5% of the *stag1a*$^{-/-}$; *stag2b*$^{-/-}$ double mutant embryos. Scale bars are 500 µm.
(B) Normal expression of *runx1* at 12 somites in *stag1a*$^{-/-}$; *stag2b*$^{-/-}$ embryos. Lateral (left) and posterior (right) views are shown. Scale bars are 500 µm. The numbers in the lower left hand corner indicate the number of embryos with the expression pattern shown.

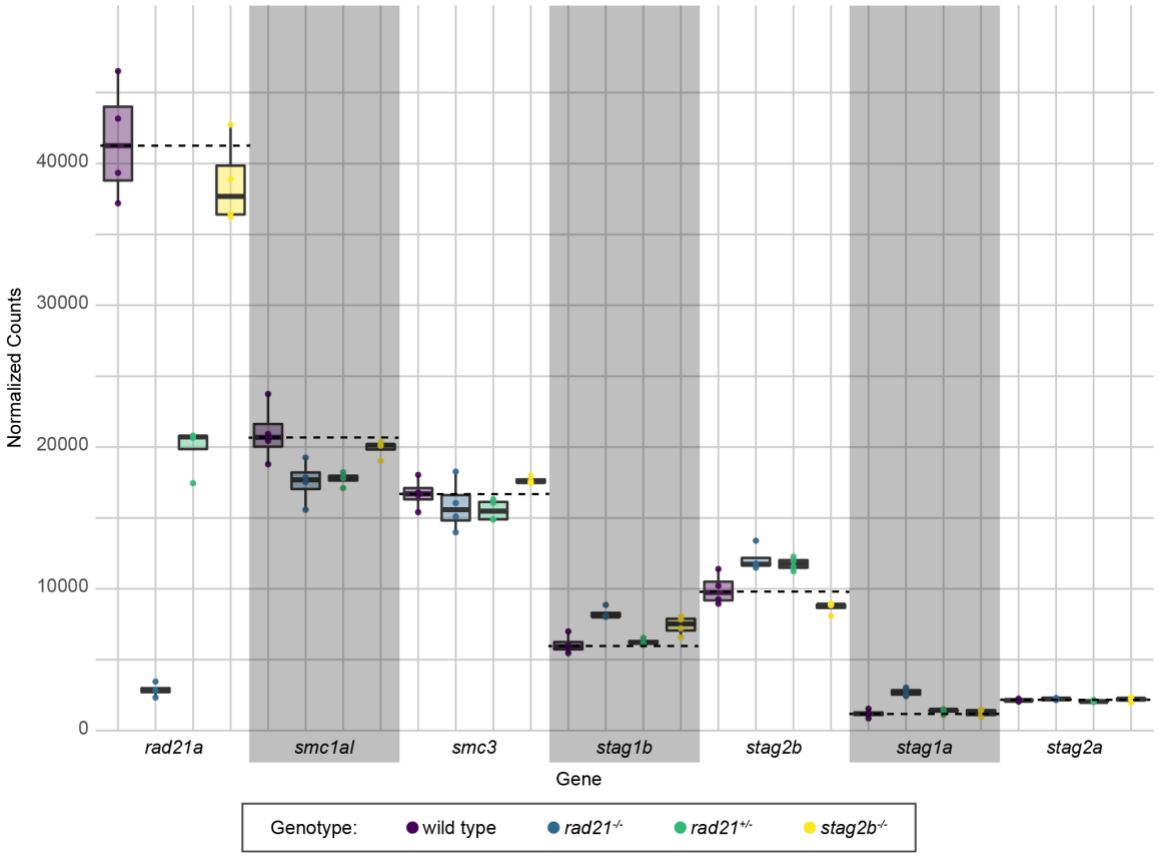

Fig. S2. Transcript counts of cohesin subunits in the tailbud.

Normalised transcript counts of *rad21a*, *smc1al*, *smc3*, *stag1b*, *stag2b*, *stag1a* and *stag2a* taken from RNA-seq of tailbuds and visualised with box plots of the 4 replicates. Genotypes are distinguished by colour: wild type samples are displayed in purple, *rad21*$^{-/-}$ in blue, *rad21*$^{+/-}$ in green, and *stag2b*$^{-/-}$ samples in yellow. For each panel, the dotted line indicates the level of expression in wild type.

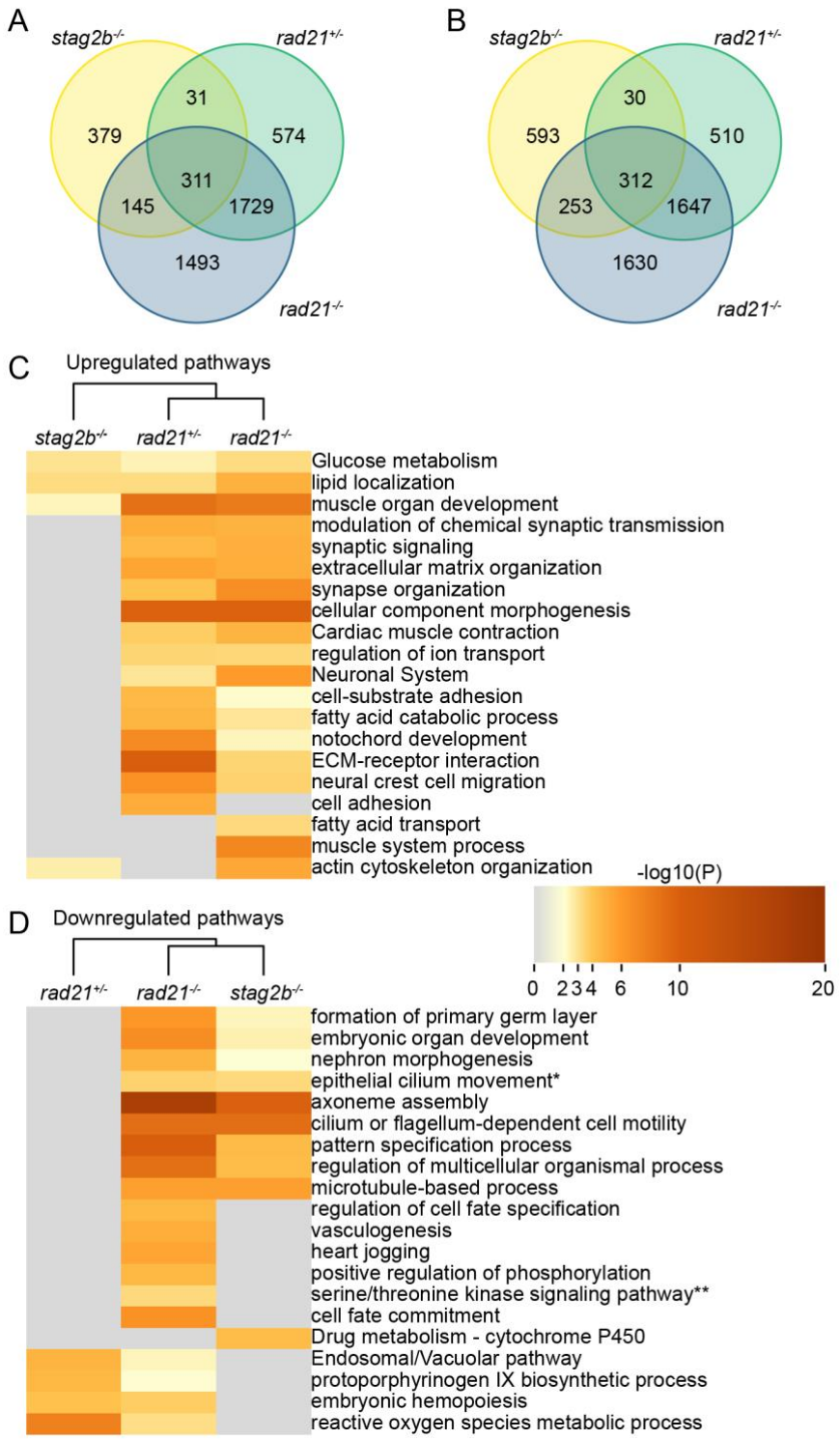

Fig. S3.

Overlap of dysregulated genes and pathway enrichment in cohesin mutant tailbuds.
(A, B) The Venn diagrams depict the overlap of significantly upregulated (A) and downregulated (B) genes in cohesin deficient tailbuds. **(C, D)** Metascape heat maps displaying the top 20 terms enriched among significantly upregulated (C) and downregulated (D) genes in cohesin-deficient tailbuds. Corresponding *p*-values are indicated by the colour scale.
*Epithelial cilium movement involved in extracellular fluid movement
**Regulation of transmembrane receptor protein serine/threonine kinase signalling pathway

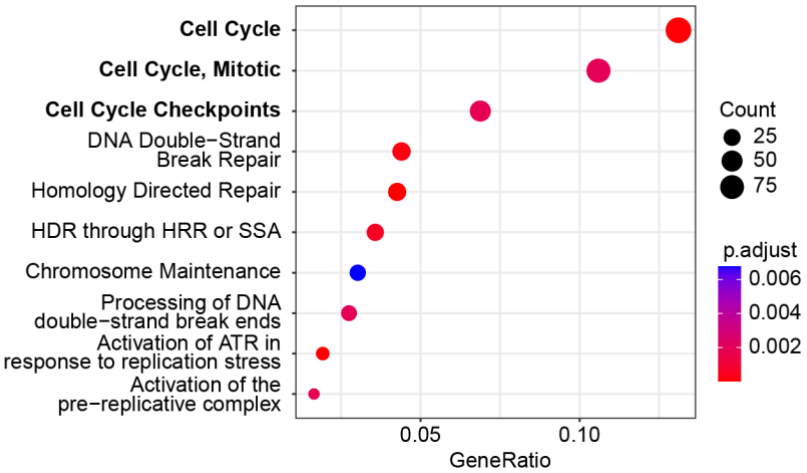

Fig. S1. Reactome analyses of downregulated pathways in *rad21*$^{-/-}$ mutant tailbuds. The dot plot shows the top 10 enriched Reactome pathways (out of 26) among the significantly downregulated genes in *rad21*$^{-/-}$ tailbuds. The size of each dot indicates the number of genes affected in the pathway, and the dot colour represents the adjusted p-value (Padj).

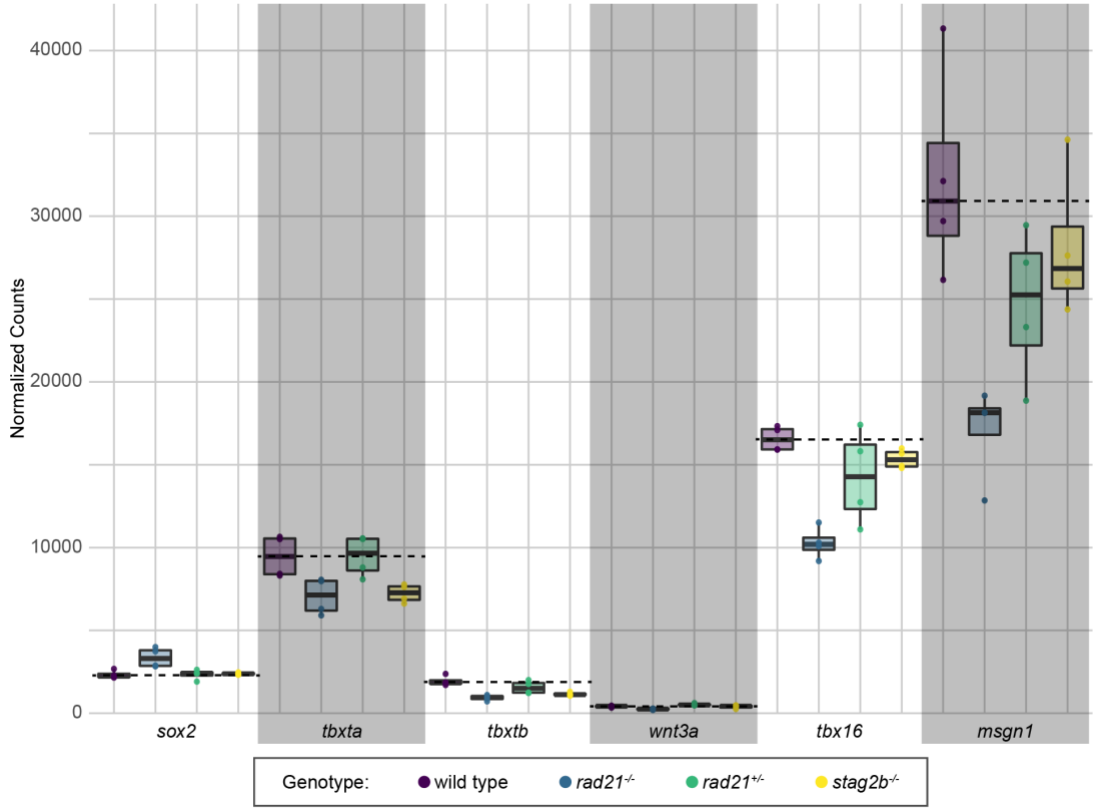

Fig. S2. Expression levels of *sox2*, *tbxta*, *tbxtb*, *wnt3a*, *tbx16* and *msgn1* in cohesin mutants. Normalised transcript counts of *sox2*, *tbxta*, *tbxtb*, *wnt3a*, *tbx16* and *msgn1* taken from tailbud RNA-seq data with 4 replicates visualised as box plots. Genotypes are distinguished by colour: wild type samples are displayed in purple, *rad21*$^{-/-}$ in blue, *rad21*$^{+/-}$ in green, and *stag2b*$^{-/-}$ samples in yellow.

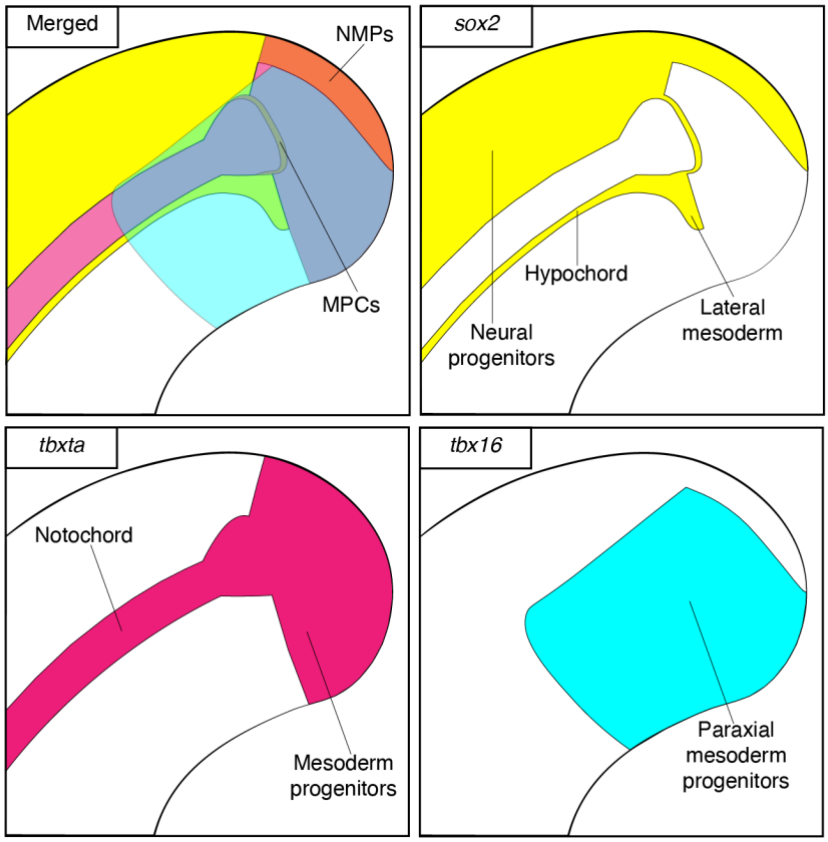

Fig. S3. Expression pattern of marker genes in the tailbud at the 16 somite stage. Schematic depicts the regions in the tailbud and progenitor types where expression of *sox2* (yellow), *tbxta* (magenta) and *tbx16* (cyan) expression is expected.

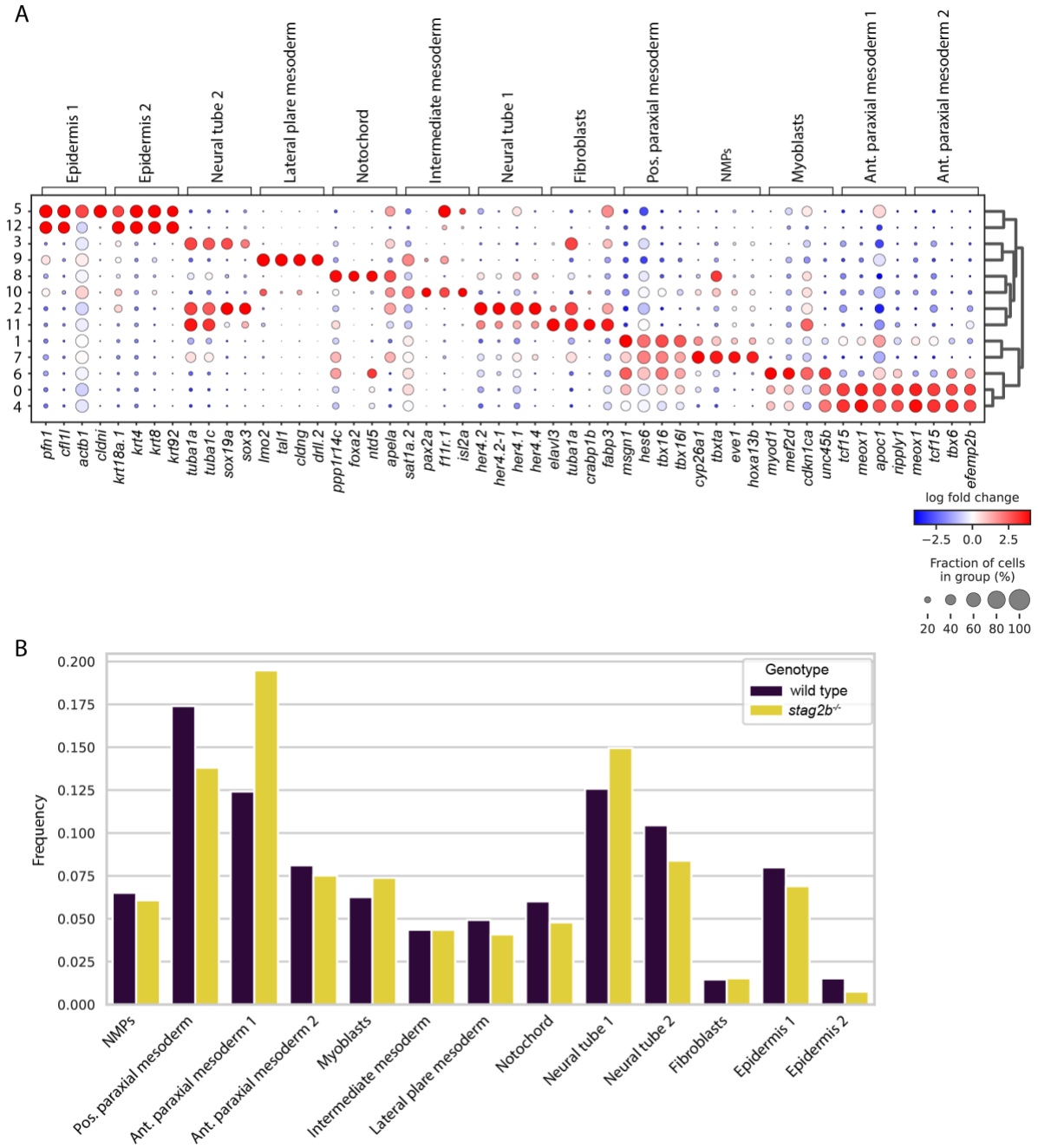

Fig. S4. Cell population analysis of single cell RNA sequencing of *stag2b*$^{-/-}$ tailbuds compared with wild type.
(A) Dot plot depicting the expression of the top 4 marker genes per cell population cluster identified (see Fig. 7A). The dot size scales with the fraction of cells expressing the gene, and the dot colour indicates the log fold change between the clusters. **(B)** Differences in the proportion of cell types between wild type and *stag2b*$^{-/-}$ tailbuds among clusters as shown in Fig. 7A.

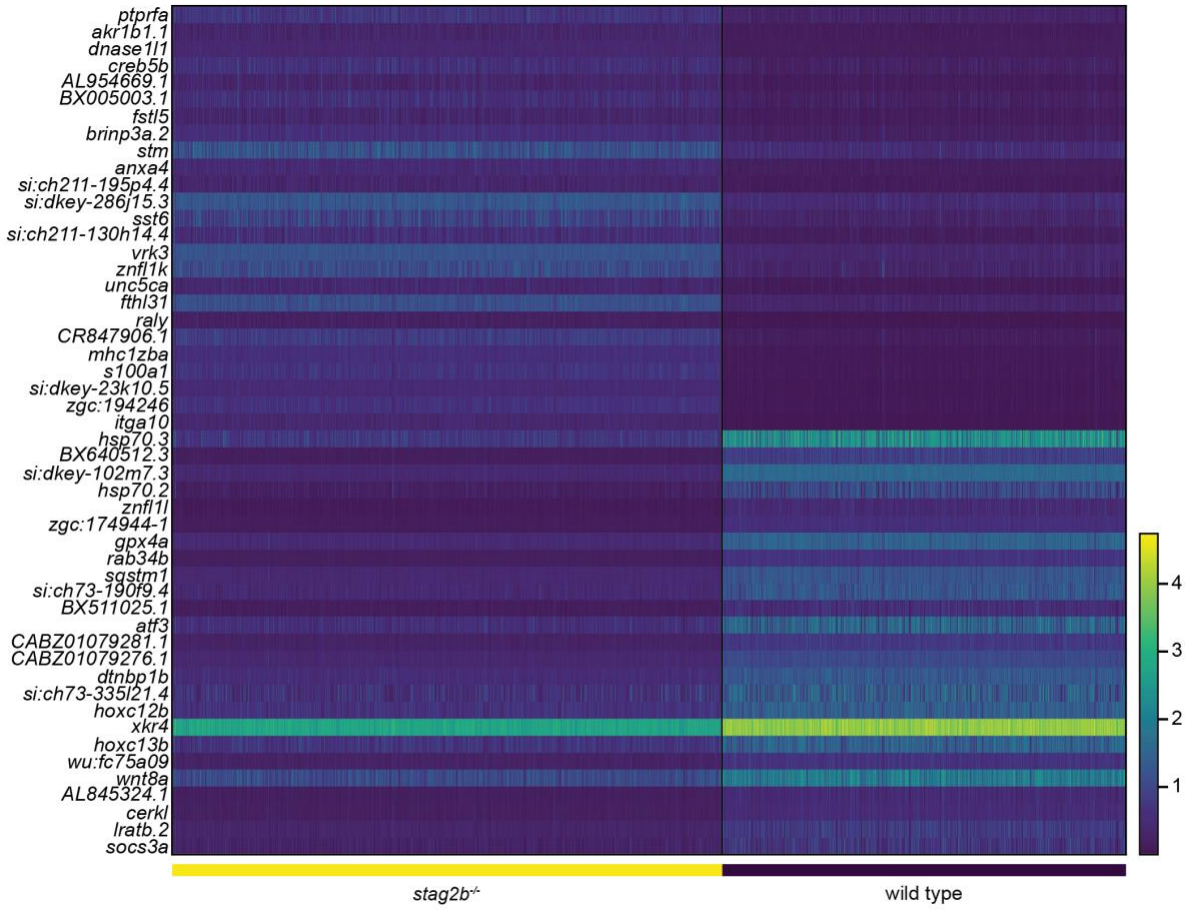

Fig. S5 . Heat map showing the top 25 differentially up- and downregulated genes between wild type and *stag2b*$^{-/-}$ NMPs. For a full list see Supplementary Data 1.

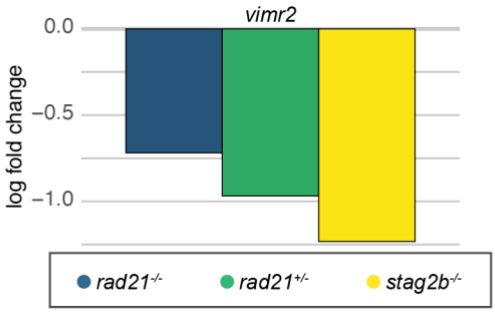

Fig. S6. *vimr2* expression in cohesin mutants. The bar graph displays the log2 fold change (5% FDR) for *vimr2* transcripts in *rad21*$^{-/-}$ (blue), *rad21*$^{+/-}$ (green) and *stag2b*$^{-/-}$ (yellow) tailbuds compared to wild type.

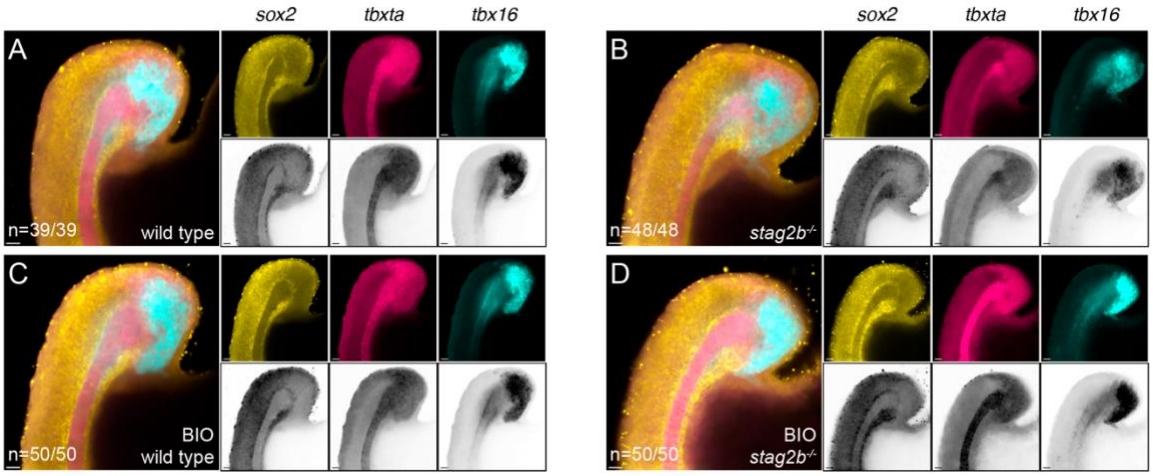

Fig. S7. Wnt stimulation rescues the notochord phenotype in *stag2b*$^{-/-}$
(A-D) Expression patterns of *sox2* (yellow), *tbxta* (magenta), and *tbx16* (cyan) in wild type (A, C) and *stag2b*$^{-/-}$ (B, D) zebrafish tailbuds with (C, D) and without (A, B) Wnt stimulation. Images are maximum intensity projections of 3 (4.8 µm) optical sections. Scale bars are 20 µm. The number of embryos with each expression pattern out of the total analysed is noted.

Data S1 (separate file): Cell cycle genes (Reactome PA) significantly (5% FDR) dysregulated in homozygous *rad21* tailbuds

Data S2 (separate file): Cell cycle genes (Reactome PA) significantly (5% FDR) dysregulated in heterozygous *rad21* tailbuds
\
Data S3 (separate file): Cell cycle genes (Reactome PA) significantly (5% FDR) dysregulated in homozygous *stag2b* tailbuds

Data S4 (separate file): Differentially expressed genes in the NMP subset comparing homozygous *stag2b* and wild type tailbuds, realated to Figure S8

Data S5 (separate file): Gene list related to canonical Wnt signalling score, related to Figure 7E

Data S6 (separate file): Wilcoxon rank sum test statistics for each cell cluster comparing the Wnt signalling score shown in Figure 7E among homozygous *stag2b* and wild type genotypes