

# Rapid and quantitative functional interrogation of human enhancer variant activity in live mice

Received: 26 December 2023

Accepted: 13 December 2024

Published online: 06 January 2025

 Check for updates

Ethan W. Hollingsworth<sup>1,2</sup>, Taryn A. Liu<sup>1</sup>, Joshua A. Alcantara<sup>1</sup>,  
Cindy X. Chen<sup>1</sup>, Sandra H. Jacinto<sup>1</sup> & Evgeny Z. Kvon<sup>1</sup>✉

Functional analysis of non-coding variants associated with congenital disorders remains challenging due to the lack of efficient in vivo models. Here we introduce dual-enSERT, a robust Cas9-based two-color fluorescent reporter system which enables rapid, quantitative comparison of enhancer allele activities in live mice in less than two weeks. We use this technology to examine and measure the gain- and loss-of-function effects of enhancer variants previously linked to limb polydactyly, autism spectrum disorder, and craniofacial malformation. By combining dual-enSERT with single-cell transcriptomics, we characterise gene expression in cells where the enhancer is normally and ectopically active, revealing candidate pathways that may lead to enhancer misregulation. Finally, we demonstrate the widespread utility of dual-enSERT by testing the effects of fifteen previously uncharacterised rare and common non-coding variants linked to neurodevelopmental disorders. In doing so we identify variants that reproducibly alter the in vivo activity of *OTX2* and *MIR9-2* brain enhancers, implicating them in autism. Dual-enSERT thus allows researchers to go from identifying candidate enhancer variants to analysis of comparative enhancer activity in live embryos in under two weeks.

The success of the large-scale genome-wide association (GWAS) and whole-genome sequencing (WGS) studies has shifted the bottleneck of human genetics from identifying sources of genetic variation to mechanistically understanding how such variation contributes to human disease<sup>1–3</sup>. Nearly 90% of disease risk-associated variation resides in non-protein coding regions of the human genome<sup>4–8</sup>. A large fraction of this variation consists of single nucleotide polymorphisms and rare variants that are hypothesised to affect transcriptional enhancers, short non-coding DNA segments that regulate cell-type-specific gene expression<sup>9–13</sup>. Each of these thousands of enhancer variants thus represents a potential entry point for understanding human disease<sup>1–3</sup>. However, the physiological effects of the vast majority of these associations remain unknown. Bridging this gap – from non-coding variants to biological mechanisms – is currently hindered by a lack of suitable in vivo

technologies for assessing if and how each human enhancer variant alters gene expression.

A major challenge is that the effects of enhancer variants on gene expression are highly cell-type-specific. For example, a typical gain-of-function enhancer variant can result in ectopic gene expression and cause pathogenic effects in cells where the enhancer is normally inactive<sup>14–21</sup>. Likewise, loss-of-function enhancer variants often result in loss of enhancer activity in one cell type, while in other cell types, its activity is unaffected<sup>22–26</sup>. These cell-type-specific effects of enhancer variants are difficult to capture with high-throughput methods such as massively parallel reporter assays (MPRAs) and CRISPR inhibitor/activator screens, both of which are primarily performed in vitro<sup>27,28</sup> or in one tissue<sup>29–34</sup>. Transgenic enhancer-reporter assays in mice enable visualisation of enhancer activity in the whole animal and are a gold standard for functionally testing when and where a human enhancer is

<sup>1</sup>Department of Developmental and Cell Biology, University of California, Irvine, CA, USA. <sup>2</sup>Medical Scientist Training Program, University of California, Irvine School of Medicine, Irvine, CA, USA. ✉ e-mail: [ekvon@uci.edu](mailto:ekvon@uci.edu)

active in vivo<sup>35–37</sup>. However, current transgenic mouse reporter assays only allow the assessment of a single enhancer per animal, precluding the direct comparison of multiple enhancer alleles. Thus, comparing activities of reference and disease-linked variant alleles requires the generation of a large number of independent transgenic mice to mitigate variation caused by mosaicism and position effects<sup>17,20,38</sup>.

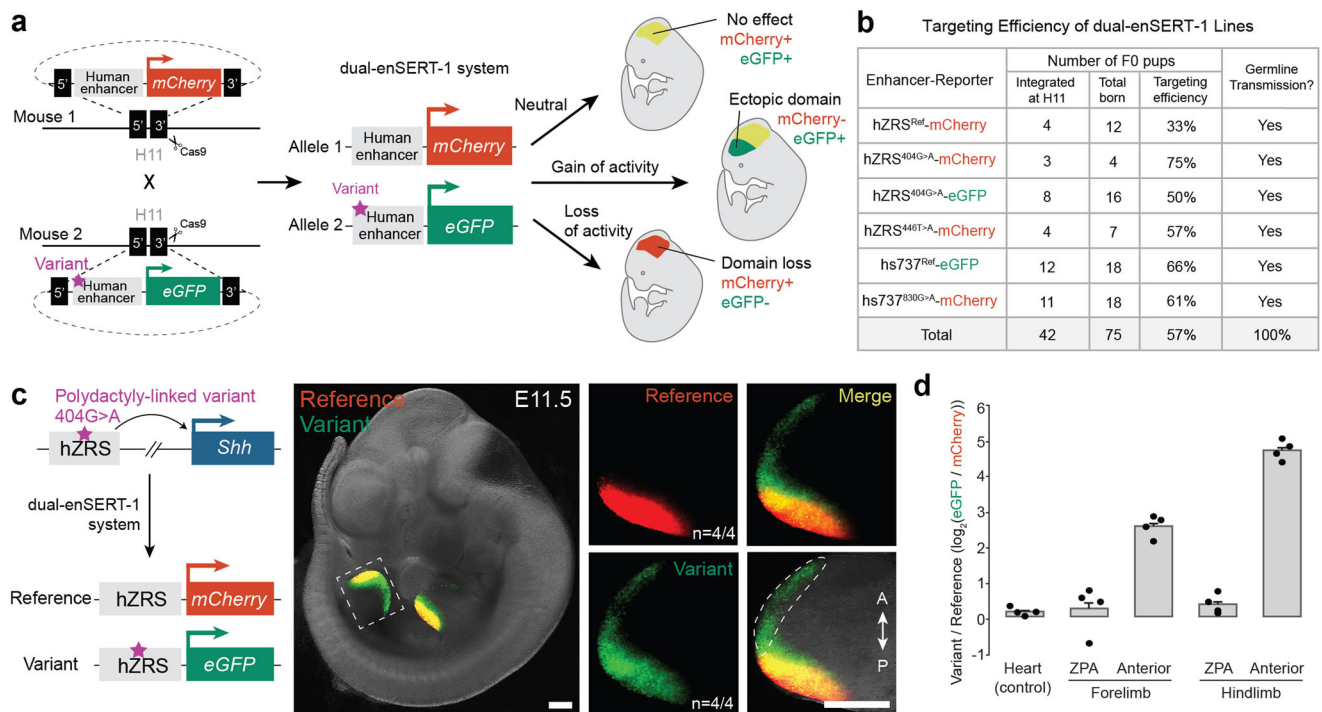
Here, we introduce dual-enSERT (dual-fluorescent enhancer inSERTion), a Cas9-based site-specific dual-fluorescent reporter system that enables simultaneous, quantitative visualisation of two human enhancer allelic activities in the same transgenic animal, thus overcoming the limitations of standard mouse reporter assays. In dual-enSERT-1, transgenes containing enhancer variants driving *eGFP* or *mCherry* are placed on different alleles of the same safe-harbour location in the mouse genome. In dual-enSERT-2, both enhancer-reporters are placed on the same transgene separated by a synthetic insulator. Using dual-enSERT-2, we were able to visualise and compare two enhancer allelic activities in live F0 mice as soon as eleven days after zygote microinjection. We first applied dual-enSERT to previously characterised pathogenic enhancer variants linked to autism spectrum disorder, limb defects, and craniofacial malformation, confirming the reported loss- and gain-of-function effects on in vivo enhancer activity. To demonstrate the utility of dual-enSERT for screening untested non-coding variants, we interrogated a panel of fifteen previously uncharacterised non-coding variants from patients with neurodevelopmental disorders and identified variants that reproducibly alter enhancer activity in vivo. Beyond the quantitative visualisation of enhancer activity, coupling dual-enSERT with single-cell transcriptomics enables the characterisation of gene expression and pathogenic enhancer activity at cellular resolution. Our dual-enSERT system is thus poised to accelerate enhancer-variant-to-function studies across many congenital disorders.

## Results

### Direct comparison of reference and variant enhancer allele activities in vivo with dual-enSERT-1

Classical mouse enhancer-reporter assays are based on random integration of the transgene into the genome<sup>35–37,39,40</sup>. Although conventional mouse transgenesis is the current gold standard for visualisation of enhancer activity in vivo, it suffers from variation due to position effects and requires the generation of a large number of transgenic animals to assess enhancer activity reproducibly<sup>17,41,42</sup>. Enhancer-reporter assays based on the integration of a transgene into a safe-harbour location of the genome overcome the problem of position effects<sup>20,43</sup>. However, due to mosaicism and inter-embryo variability, these site-specific assays also require analysis of a large number of transgenic animals, especially for detecting the subtle effects of enhancer mutations<sup>20,38,44</sup>. To overcome these limitations, we developed dual-enSERT-1, a transgenic approach based on highly efficient Cas9-mediated integration of enhancers driving fluorescent reporters into the H11 safe-harbour integration site<sup>45</sup>. One enhancer allele is placed upstream of an *eGFP* reporter and the second enhancer allele is placed upstream of a *mCherry* reporter followed by Cas9-mediated integration of each transgene into the H11 locus (Fig. 1a). With dual-enSERT-1, we achieved an average transgenic targeting efficiency of 57% across all single-reporter constructs tested in this study, and all insertions were germline transmissible (Fig. 1b).

We first assessed whether we could detect and quantify the effects of non-coding variants on enhancer activity with dual-enSERT-1 by testing a previously characterised pathogenic allele of the ZRS (zone of polarising activity) Regulatory Sequence, also known as MFCS1) enhancer of *Sonic hedgehog* (*Shh*). Single nucleotide variants in the ZRS cause congenital limb malformations, most typically preaxial polydactyly, in humans, cats, chickens, and mice<sup>20,46,47</sup>. ZRS variants



**Fig. 1 | Simultaneous comparison of human reference and variant enhancer activities in developing mouse embryos.** **a** Schematic overview of the dual-enSERT-1 strategy and its potential readouts for different types of enhancer variants. **b** Transgenic targeting efficiencies of all dual-enSERT-1 constructs generated in this study. **c** Representative images of transgenic hZRS<sup>Ref</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP embryos at E11.5. A close-up of the hindlimb with separate channels and an outline depicting the gain of enhancer activity (green channel) in

the anterior domain are shown. A, anterior; P, posterior. **d** Plots quantifying fold-change (log<sub>2</sub>) difference in normalised reporter fluorescence between variant and reference hZRS alleles. Two-sided paired *t* tests vs. Heart: Forelimb ZPA, *P* = ns; Forelimb Anterior, *P* = 0.00072; Hindlimb ZPA, *P* = ns; Hindlimb Anterior, *P* = 9.48E-08. Data represented as mean ± SEM. Data points represent independent biological replicates (*n* = 4 embryos). Scale bars, 500 μm. Source data are provided as a Source Data file.

implicated in preaxial polydactyly cause ectopic *Shh* expression in the anterior portion of the developing limb bud, leading to erroneous digit outgrowth similar to human patients<sup>20,47</sup>. We created two stable transgenic mouse lines, one with the human reference ZRS allele driving *mCherry* (hZRS<sup>ref</sup>-*mCherry*) and a second line with a previously characterised pathogenic hZRS allele containing the polydactyly-linked 404G>A variant driving *eGFP* (hZRS<sup>404G>A</sup>-*eGFP*) (Fig. 1c). To visualise reference and variant hZRS enhancer activities simultaneously, we crossed these mouse lines to generate two-colour dual-enSERT-1 embryos. Both lines had two copies of respective transgene integrated at the HII locus enabling direct and quantitative comparison between enhancer alleles in two-colour dual-enSERT-1 embryos (Supplementary Fig. 1a, b and “Methods”). In these hZRS<sup>ref</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP* transgenic embryos, mCherry fluorescence was detected in the ZPA of fore- and hindlimb buds, matching the location of normal *Shh* expression ( $n = 4/4$  embryos)<sup>48</sup>. We also detected weaker mCherry fluorescence elsewhere in the embryo, including the heart, consistent with the weak endogenous activity of the *Hsp68* promoter<sup>20</sup>. EGFP expression pattern driven by the hZRS<sup>404G>A</sup> variant allele was indistinguishable from the mCherry expression pattern except for the anterior limb region. eGFP fluorescence extended into the anterior domain of the limb bud in all examined embryos, mimicking ectopic *Shh* misexpression in mice with polydactyly ( $n = 4/4$ ; Fig. 1c, d).

To quantify the effect of 404G>A variant on hZRS enhancer activity, we compared eGFP and mCherry fluorescent intensities within the same embryo. We detected similar levels of eGFP and mCherry fluorescence in the heart (1.1-fold difference, eGFP vs. mCherry,  $P = ns$ ), indicating that differences in eGFP and mCherry maturation times and half-lives have a negligible effect on our measurements. Nevertheless, to exclude even small confounding effects caused by differences in the choice of the reporter, we used promoter-driven heart fluorescence as an endogenous control for all future comparisons (“Methods”). Both alleles drove similar levels of reporter expression in the ZPA (Forelimb,  $P = ns$ ; Hindlimb,  $P = ns$ ). We detected a 6.5-fold stronger reporter expression in the anterior forelimb bud ( $P = 0.00072$ ) and a 31-fold stronger expression in the anterior hindlimb bud ( $P = 9.84E-08$ ). These results indicate that dual-enSERT-1 can robustly detect and quantify changes in limb enhancer activity caused by pathogenic hZRS variants.

We next asked whether dual-enSERT-1 could be used to study human non-coding variants linked to other congenital disorders. We focused on the hs737 enhancer of *EBF3*, a region where several independent rare variants have been identified in patients with autism spectrum disorder and intellectual disability<sup>17,49</sup>. We generated a transgenic mouse line in which the human reference hs737 allele drives *eGFP* (hs737<sup>ref</sup>-*eGFP*) and a second line in which an 830G>A variant allele, identified in a patient with autism, drives *mCherry* (hs737<sup>830G>A</sup>-*mCherry*) (Supplementary Fig. 1c). We then bred these mouse lines each containing single-copy integrated transgenes and examined reporter gene expression in E11.5 embryos (Supplementary Fig. 1a, b). Live imaging revealed comparable levels of eGFP and mCherry fluorescence in the midbrain, hindbrain, and neural tube (Midbrain,  $P = ns$ ; Hindbrain,  $P = ns$ ; Neural Tube,  $P = ns$ ; Supplementary Fig. 1c, d). However, mCherry expression driven by the hs737<sup>830G>A</sup> variant allele also extended into the forebrain in all examined hs737<sup>ref</sup>-*eGFP*/hs737<sup>830G>A</sup>-*mCherry* embryos (9.4-fold difference in Forebrain,  $P = 0.0043$ ; Supplementary Fig. 1c, d). These results are consistent with previous observations of ectopic forebrain activity using non-quantitative LacZ-based transgenic assays<sup>17</sup>.

### Comparative functional assessment of independent enhancer variants

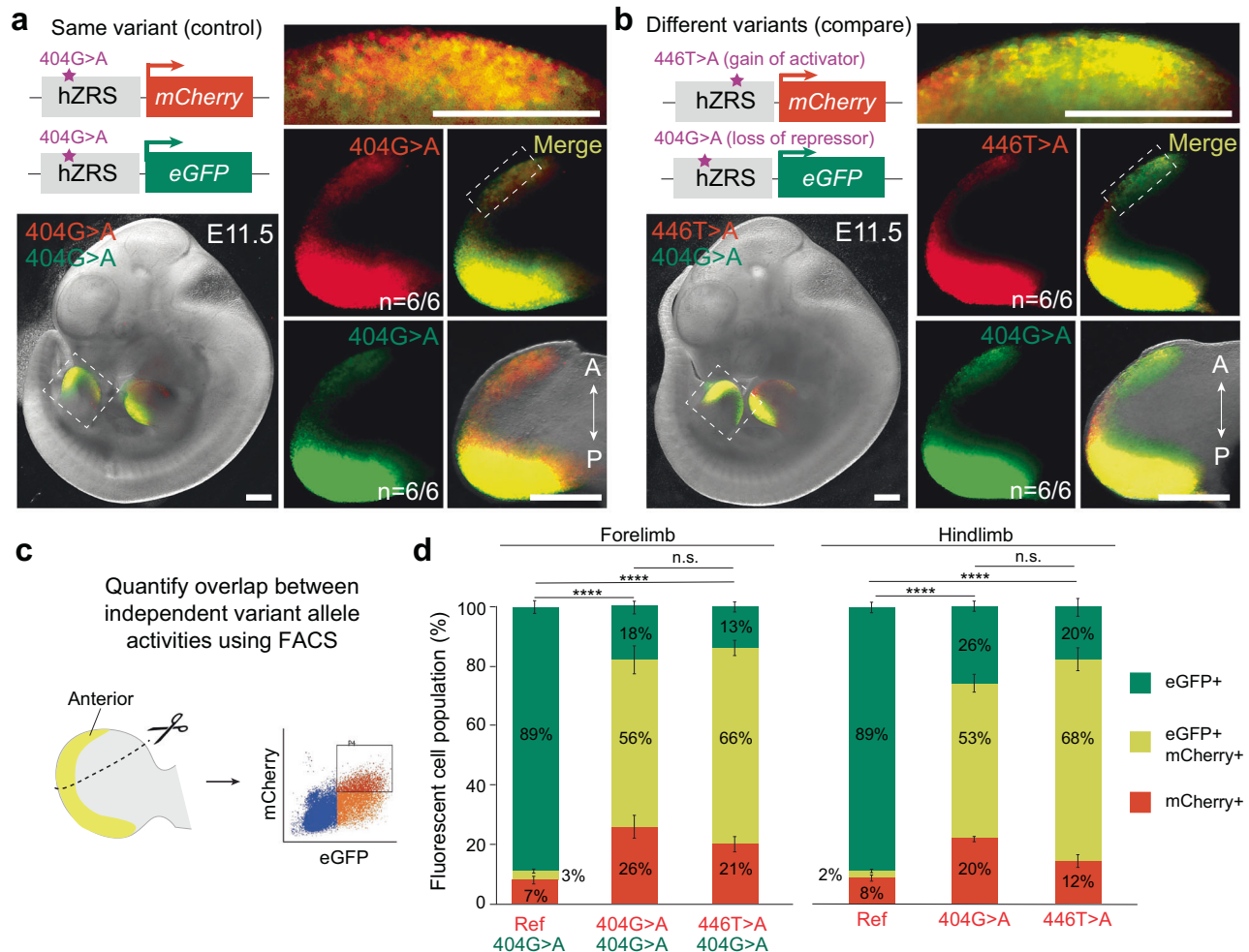
We next tested the ability of dual-enSERT-1 to simultaneously visualise and compare the effects of independent enhancer variants in live mice. Human genetics studies often identify disease-linked hotspots in which multiple rare variants affect the same enhancer<sup>11,20,50–53</sup>. For

example, 22 different rare human point mutations in the hZRS enhancer have been identified in patients with polydactyly<sup>20,47</sup>. Despite extensive work on this enhancer, it is unknown if these independent mutations result in ectopic gene expression in the same or different cell populations of the limb bud.

As a positive control, we first created a mouse line in which hZRS<sup>404G>A</sup> allele drives *mCherry* and crossed this line to the mouse line in which the same hZRS<sup>404G>A</sup> allele drives *eGFP*. We collected the resulting two-colour hZRS<sup>404G>A</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP* transgenic embryos with the expectation that most fluorescent cells will be double positive eGFP+/mCherry+ cells (Fig. 2a and Supplementary Fig. 2a). Indeed, at E11.5, the overlap in ectopic GFP and mCherry activity in the anterior limb bud mesenchyme was visually indistinguishable (Fig. 2a and Supplementary Fig. 2a). To quantify this overlap at cellular resolution, we used fluorescence-activated cell sorting (FACS) to isolate double-positive cells from anterior limb bud mesenchyme (Fig. 2c). As a negative control, we used transgenic mice in which the ZRS<sup>404G>A</sup> variant allele was driving *eGFP* and the ZRS reference allele was driving *mCherry* (hZRS<sup>ref</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP*) with the expectation that only eGFP+ cells should be present in the anterior limb bud cell population (Fig. 1c). Indeed, 92% of fluorescent cells sorted from anterior limb buds of hZRS<sup>ref</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP* embryos were eGFP+/mCherry- and only 3% were eGFP+/mCherry+, confirming allele-specific ectopic expression of the ZRS<sup>404G>A</sup> allele (Fig. 2c, d and Supplementary Fig. 3a, b). We next examined hZRS<sup>404G>A</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP* transgenic embryos carrying the ZRS<sup>404G>A</sup> variant allele driving both colours. Only 56% (in forelimbs) to 53% (in hindlimbs) of fluorescent cells in anterior limb buds were eGFP+/mCherry+ (Fig. 2d and Supplementary Fig. 3c).

To test if this variability could be caused by differences between fluorophores, we performed *mCherry* and *eGFP* mRNA quantification in different populations of anterior limb bud cells. We dissected the anterior domains of hindlimbs from hZRS<sup>404G>A</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP* embryos and sorted them into four cell populations: mCherry+/eGFP+, mCherry+/eGFP-, mCherry-/eGFP+, and mCherry-/eGFP-. We then performed qPCR to quantify *mCherry* and *eGFP* mRNA levels in each of these cell populations (Supplementary Fig. 3d). Both eGFP-positive cell populations (mCherry+/eGFP+ and mCherry-/eGFP+) expressed more than 8-fold higher levels of *eGFP* than eGFP-negative cell populations (mCherry+/eGFP- vs mCherry-/eGFP-,  $P = 0.008$ ; mCherry-/eGFP+ vs mCherry+/eGFP-,  $P = 0.009$ ) (Supplementary Fig. 3d). This indicates that eGFP fluorescence accurately reflects *eGFP* expression. Both mCherry-positive cell populations displayed comparable levels of *eGFP* and *mCherry* transcripts ( $P = NS$ ). However, similar levels of *mCherry* transcripts were also observed in mCherry-negative cells ( $P = NS$ ) (Supplementary Fig. 3d). The incomplete fluorescent overlap on a single-cell level is likely due to post-transcriptional differences between fluorophores. For example, a significantly longer maturation time of mCherry in comparison to eGFP is consistent with eGFP+ cells expressing *mCherry* transcripts, but not mature protein<sup>54,55</sup>. We mitigated the effect of this variability on our measurements of enhancer activity by quantifying fluorescence over the entire population of cells and using the heart as an endogenous control (Fig. 1d).

We next generated a transgenic mouse line in which a hZRS<sup>446T>A</sup> variant allele identified in a family with preaxial polydactyly drives *mCherry* (hZRS<sup>446T>A</sup>-*mCherry*)<sup>56</sup>. The 446T>A variant is hypothesised to create a de novo activator binding site, but its effect on in vivo hZRS enhancer activity is unknown. In contrast, the well-characterised 404G>A variant disrupts a repressor binding site and causes ectopic reporter expression in the anterior limb bud mesenchyme<sup>41,47</sup>. To visualise hZRS<sup>404G>A</sup> and hZRS<sup>446T>A</sup> variant allele activities simultaneously, we bred these mouse lines, each containing two copies of the respective transgene to generate two-colour transgenic embryos (Supplementary Fig. 1a, b). In this hZRS<sup>446T>A</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP* transgenic embryos, mCherry and eGFP were detected in a highly



**Fig. 2 | Comparison of the effects of independent human-disease-linked variants in the hZRS enhancer.** **a** Sample image of an E11.5 hZRS<sup>404G>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP embryo. Panels on the right show high-resolution images of the hindlimb bud and anterior domain (above), as marked by outlined boxes. **b** Representative image of hZRS<sup>446T>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP embryo at E11.5. Panels on the right show high-resolution images of the hindlimb bud and its anterior domain (above), as marked by outlined boxes. **c** FACS-based quantification of the overlap between mCherry- and eGFP-expressing cells in the anterior domain of limb buds. **d** Plots depicting population distribution of eGFP+, eGFP+mCherry+, and mCherry+ cells in the anterior domain of fore- and hindlimbs. Each

genotype contains at least three independent biological replicates. Genotypes of dual-enSERT-1 embryos are labelled along the x-axis and coloured according to the downstream reporter gene. Fisher's exact tests. hZRS<sup>Ref</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP vs. hZRS<sup>404G>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP: Forelimb,  $P < 2.2E-16$ ; Hindlimb,  $P < 2.2E-16$ . hZRS<sup>Ref</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP vs. hZRS<sup>446T>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP: Forelimb,  $P < 2.2E-16$ ; Hindlimb,  $P < 2.2E-16$ . hZRS<sup>404G>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP vs. hZRS<sup>446T>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP: Forelimb,  $P = ns$ ; Hindlimb,  $P = ns$ . Data represented as mean  $\pm$  SEM for plots. All scale bars, 500  $\mu$ m. Source data are provided as a Source Data file.

overlapping pattern in the ZPA and anterior limb bud mesenchyme at E11.5 (ZPA,  $P = ns$ ; Anterior,  $P = ns$ ; Fig. 2b and Supplementary Fig. 2b).

The overlap in ectopic activity in the anterior limb bud mesenchyme was visually and quantitatively indistinguishable from the overlap observed in hZRS<sup>404G>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP transgenic embryos in which eGFP and mCherry were driven by the same hZRS<sup>404G>A</sup> variant allele (Fig. 2a, b and Supplementary Fig. 2a). We next examined this extent of overlap at cellular resolution in anterior cells from hZRS<sup>446T>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP transgenic embryos in which eGFP and mCherry were driven by different hZRS variants (Fig. 2b and Supplementary Fig. 3e). 66% (forelimb) to 68% (in hindlimb) of fluorescent cells in anterior limb buds were eGFP+/mCherry+ (Fig. 2d). This fraction of double-positive anterior limb bud cells was not significantly different from the fraction of double-positive cells in hZRS<sup>404G>A</sup>-mCherry/hZRS<sup>404G>A</sup>-eGFP transgenic embryos. These data indicate that the 404G>A variant, which disrupts a repressor binding site, and the 446T>A variant, which creates an activator binding site,

both cause highly overlapping ectopic expression in the same population of anterior limb cells.

### Dual-enSERT-2 allows rapid comparison of enhancer allele activities in FO mice

A limitation of the dual-enSERT-1 system is the ~6-month time required to obtain two-colour F2 embryos. This protracted timeframe limits the number of enhancer variants that can be rapidly tested in mice. To overcome this bottleneck, we constructed single transgenes containing mCherry and eGFP reporters driven by different enhancer alleles in divergent orientations. With this bicistronic system, henceforth referred to as dual-enSERT-2, an injection of a single construct would yield two-colour FO embryos in as little as eleven days (Supplementary Fig. 4a). As a proof of principle, we placed the hZRS<sup>Ref</sup> allele upstream of mCherry and hZRS<sup>404G>A</sup> variant allele upstream of eGFP all within the same construct. To prevent cross-activation between enhancer alleles and reporter genes, we separated two transgenes with three copies of

the well-characterised chicken  $\beta$ -globin insulator 5'-HS4. 5'-HS4 is widely used for its robust ability to block enhancer-promoter activation in the genome<sup>57–59</sup> and in the context of a zebrafish transgene<sup>24</sup>. To our surprise, separating the transgenes with three copies of 5'-HS4 completely failed to prevent enhancer cross-activation in single- and multiple-copy transgenes (Supplementary Fig. 4b, c and “Methods”).

We thought to create a stronger synthetic insulator (SI) consisting of A2 (two copies), ALOXE3, and 5'-HS4 (two copies) insulators and placed it between two transgenes as well as in the vector backbone to prevent cross-activation between transgene copies<sup>57,60,61</sup> (“Methods”). In embryos with a single-copy transgene integration of hZRS<sup>ref</sup>-mCherry/SI/hZRS<sup>404G>A</sup>-eGFP at the H11 locus, mCherry fluorescence was restricted to the ZPA while eGFP fluorescence was also observed in the anterior limb bud (3/3 embryos; forelimb and hindlimb ZPA,  $P = ns$ ; 1.9-fold in anterior forelimb,  $P = 0.0229$ ; 2.1-fold in anterior hindlimb,  $P = 0.00898$ ; Supplementary Fig. 4d, e). These results mimic those of dual-enSERT-1 in which enhancer-reporters were placed on separate H11 alleles and indicate that the SI can prevent reporter cross-activation in the anterior limb bud. Furthermore, in transgenic embryos containing a single copy of the mouse ZRS (mZRS) driving mCherry and an enhancerless eGFP transgene (mZRS<sup>ref</sup>-mCherry/SI/empty-eGFP), we observed only mCherry fluorescence in E11.5 limb buds, but no detectable eGFP fluorescence, indicating that the SI can fully insulate the two transgenes (Supplementary Fig. S4h, i)<sup>62</sup>.

Interestingly, in embryos with multiple copies of a transgene separated by a synthetic insulator at the H11 locus, both anterior and posterior limb buds showed robust mCherry and eGFP expression in all examined embryos (8/8 embryos; Forelimb ZPA,  $P = ns$ ; Forelimb Anterior,  $P = ns$ ; Hindlimb ZPA,  $P = ns$ ; Hindlimb Anterior,  $P = ns$ ; Supplementary Fig. 4f, g). These results indicate that the hZRS can bypass the synthetic insulator in the context of a multi-copy transgene with multiple enhancer-reporters flanked by synthetic insulators (Supplementary Fig. 4j). Therefore, a synthetic insulator-based dual-enSERT-2 can discriminate between enhancer allele activities only if a single copy of the bicistronic transgene is integrated at the H11 landing site (Supplementary Fig. 4j).

To optimise the efficiency of dual-enSERT-2, we sought to maximise the number of single-copy integrants at the H11 landing site. Recent work in zebrafish and mice has shown that the addition of biotinylated nucleotides to the ends of donor DNA prevents concatemer formation during Cas9-mediated homology-directed repair<sup>63,64</sup>. To test if the addition of biotin (B) results in preferential single-copy transgene integration at the H11 locus, we added biotinylated nucleotides to the ends of a linearised FO dual-enSERT vector carrying the human reference and the 404G>A hZRS variant alleles (“Methods”). We injected this B-hZRS<sup>ref</sup>-mCherry/SI/hZRS<sup>404G>A</sup>-eGFP-B construct into mouse zygotes together with Cas9 ribonucleoproteins (RNPs) and imaged the FO embryos eleven days later. mCherry fluorescence was restricted to the ZPA, while hZRS<sup>404G>A</sup>-driven eGFP expression extended into the anterior limb bud in all embryos with transgene integration at the H11 locus (5/5 embryos, Forelimb ZPA,  $P = ns$ ; 1.4-fold difference in Forelimb Anterior,  $P = 0.012$ ; Hindlimb ZPA,  $P = ns$ ; 1.6-fold difference in Hindlimb Anterior,  $P = 0.00037$ ; Fig. 3a, b and Supplementary Data File 1). Moreover, genotyping confirmed that most transgenic embryos ( $n = 5/6$  embryos) contained a single-copy transgene at the H11 locus and lacked concatemers (Fig. 3a, b and Supplementary Data File 1). This result indicates that the addition of biotinylated nucleotides results in preferential single-copy integration of a bicistronic transgene and enables discrimination between enhancer allele activities in a straightforward manner.

### Dual-enSERT-2 detects the effects of different types of pathogenic enhancer variants

To test if the optimised dual-enSERT-2 can rapidly detect the qualitative effects of other disease-linked non-coding variants, we returned

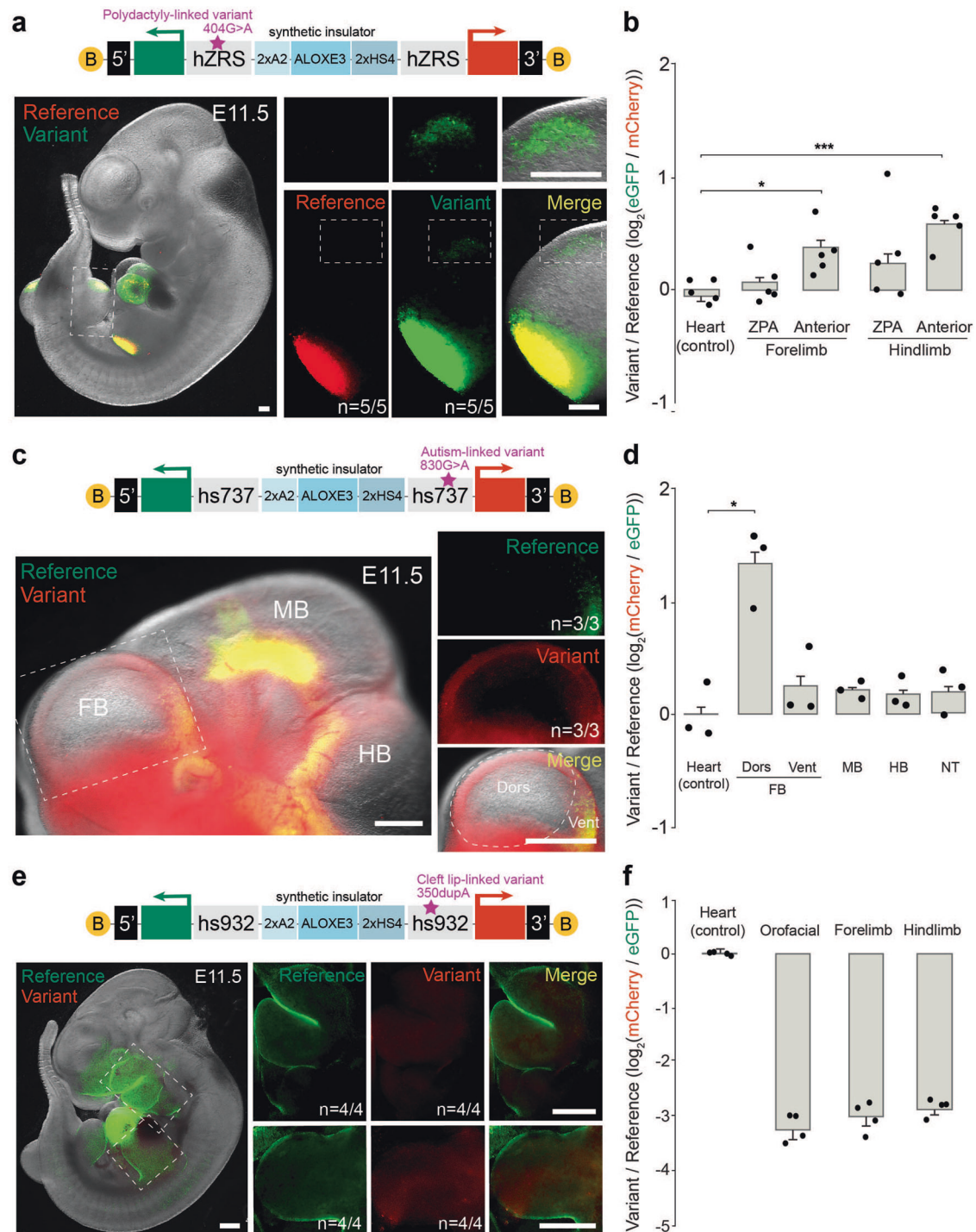
to the autism- and intellectual disability-linked *hs737/EBF3* locus (Supplementary Fig. 1c, d). We placed the *hs737*<sup>ref</sup> allele upstream of *eGFP* and the *hs737*<sup>830G>A</sup> variant allele upstream of *mCherry*. We then linearised, biotinylated, and injected the resulting B-*hs737*<sup>ref</sup>-*eGFP*/SI/*hs737*<sup>830G>A</sup>-*mCherry*-B construct into mouse zygotes with Cas9 RNPs. Eleven days later, live imaging revealed increased mCherry reporter expression in the dorsal forebrain (1.3-fold,  $P = 0.044$ ) while the ventral forebrain ( $P = ns$ ), midbrain ( $P = ns$ ), hindbrain ( $P = ns$ ), and neural tube ( $P = ns$ ) showed no difference between eGFP and mCherry reporter expression (Fig. 3c d, and Supplementary Data File 1). These results reproduce our earlier results using dual-enSERT-1, but in a much shorter time frame (11 days vs. 6 months).

We next asked if dual-enSERT-2 could be employed to study human disease variants that cause loss of enhancer activity. We focused on the previously characterised rare non-coding 350dupA mutation at the *IRF6* locus that is linked to cleft lip formation<sup>42</sup>. 350dupA is a single A duplication at position 350 of the *hs932* face enhancer of *IRF6* (also known as MCS9.7). We created a bicistronic vector with the *hs932* reference allele driving *eGFP* and the *hs932*<sup>350dupA</sup> variant allele driving *mCherry* separated by the synthetic insulator (B-*hs932*<sup>ref</sup>-*eGFP*/SI/*hs932*<sup>350dupA</sup>-*mCherry*-B; Fig. 3e). Imaging of transgenic mouse embryos at E11.5 revealed strong eGFP expression driven by the reference *hs932* allele in the orofacial and limb ectoderm, as previously reported using LacZ-based transgenesis<sup>42</sup> (4/4 embryos, Fig. 3f and Supplementary Data File 1). By contrast, we found no mCherry fluorescence in all examined transgenic embryos, indicating a near complete loss of enhancer activity (10-fold in orofacial,  $P = 0.00051$ ; 8.5-fold in forelimb,  $P = 0.00015$ ; 7.7-fold in hindlimb,  $P = 0.00022$ ; Fig. 3e, f). Overall, 80% of all FO mice with reporter integration at the H11 locus generated with biotinylated dual-enSERT-2 constructs contained only single-copy transgenes (Supplementary Data File 1). Altogether, these results demonstrate the utility of dual-enSERT-2 for rapidly detecting and quantifying gain- and loss-of activity for disease-linked enhancer variants in vivo.

### Scaled assessment of previously uncharacterised non-coding variants with dual-enSERT-2

We next sought to use dual-enSERT-2 to functionally screen previously uncharacterised human non-coding variants linked to congenital disease. We compiled a list of thousands of candidate pathogenic rare and common non-coding variants from patients with neurodevelopmental disorders (NDDs) identified by GWAS and WGS studies<sup>17,26,65–68</sup>. We intersected this list with previously validated in vivo human and mouse enhancers active at embryonic day E11.5. We further narrowed the list by focusing on enhancers affected by multiple independent non-coding variants or enhancers with a known NDD-linked target gene based on capture Hi-C data, reasoning these regions the least likely to be random mutations (Fig. 4a, Table 1 and “Methods”) <sup>69</sup>. From this prioritisation, we chose fourteen rare SNVs and one common indel distributed between seven unique enhancers active in the forebrain, midbrain, hindbrain, or neural tube.

To efficiently test these prioritised variants, we generated compound variant alleles for each of the enhancers, in some instances up to three independent variants per enhancer (Fig. 4a). We placed these compound variant alleles upstream of *eGFP* and the corresponding reference enhancer alleles upstream of *mCherry* and compared them using dual-enSERT-2. We observed loss of enhancer activity in one enhancer (*hs268*), the gain of enhancer activity for one enhancer (*hs1791*), no detectable changes for four enhancers (Fig. 4 and Supplementary Fig. 5a–h), and one enhancer (*hMM1518*) was inactive in our assay possibly due to interspecies sequence divergence (Supplementary Fig. 5i). For example, the introduction of two rare, autism-linked 435G>T and 700C>T variants in the *hs268* enhancer of *MIR9-2* resulted in substantial loss of enhancer activity in the brain and neural



**Fig. 3 | An F0-based dual-enSERT-2 system for rapid testing of human enhancer variant activity.** **a** Fluorescent image of a B-hZRS<sup>ref</sup>-mCherry/SI/hZRS<sup>404G>A</sup>-eGFP-B whole embryo at E11.5 with close-up images of the hindlimb and anterior domain on right as marked by white dotted boxes. B, biotin. Scale bars, 250  $\mu$ m. **b** Plots quantifying fold-change ( $\log_2$ ) difference in reporter intensity between variant and reference alleles by tissue in B-hZRS<sup>ref</sup>-mCherry/SI/hZRS<sup>404G>A</sup>-eGFP-B embryos. Data points represent independent biological replicates ( $n = 5$  embryos). Two-sided paired  $t$  tests vs. Heart: Forelimb ZPA,  $P = ns$ ; Forelimb Anterior,  $P = 0.0124$ ; Hindlimb ZPA,  $P = ns$ ; Hindlimb Anterior,  $P = 0.00037$ . Data represented as mean  $\pm$  SEM. **c** Sample fluorescent image of a B-hs737<sup>ref</sup>-eGFP/SI/h737<sup>830G>A</sup>-mCherry-B embryo at E11.5 with high-resolution image of forebrain on right. The white outline highlights ectopic areas with the gain in enhancer activity (red channel). B, biotin. Scale bars, 500  $\mu$ m. **d** Plot depicting quantification of fold-change ( $\log_2$ ) difference in variant-reference reporter intensity by tissue in

B-hs737<sup>ref</sup>-eGFP/SI/h737<sup>830G>A</sup>-mCherry-B embryos. Data points represent independent biological replicates ( $n = 3$  embryos). Two-sided paired  $t$ -tests vs. Heart: Dorsal (Dors) Forebrain,  $P = 0.044$ ; Ventral (Vent) Forebrain (FB),  $P = ns$ ; Midbrain (MB),  $P = ns$ ; Hindbrain (HB),  $P = ns$ ; Neural Tube (NT),  $P = ns$ . Data represented as mean  $\pm$  SEM. **e** Representative image of a B-hs932<sup>350dupA</sup>-mCherry/SI/h932<sup>ref</sup>-eGFP-B embryo, with expanded views on the right to highlight the orofacial region and limbs at E11.5, as marked by white dotted boxes. B, biotin. Scale bars, 500  $\mu$ m. **f** Quantitative plot for fold-change ( $\log_2$ ) difference in variant-reference fluorescent reporter intensity for heart, orofacial region, and limbs from B-hs932<sup>ref</sup>-eGFP/SI/h932<sup>350dupA</sup>-mCherry-B embryos. Data points represent four biological replicates ( $n = 4$  embryos). Two-sided paired  $t$  tests vs. Heart: Orofacial,  $P = 0.00051$ , Forelimb,  $P = 0.00015$ ; Hindlimb,  $P = 0.00022$ . Data represented as mean  $\pm$  SEM. Source data are provided as a Source Data file.

tube (5/5 embryos; 4.4-fold in the forebrain,  $P=0.001$ ; 3.2-fold in the midbrain,  $P=0.0093$ ; 3.3-fold in the hindbrain,  $P=0.0065$ ; 3.4-fold in the neural tube,  $P=0.0056$ ; Fig. 4b, c). 435G>T and 700C>T variants disrupt evolutionary conserved putative binding sites for neuronally-expressed transcription factors TBX/TBR2 and BCL11A, respectively (Fig. 4d). Conversely, a variant allele of the hs1791 midbrain enhancer of *OTX2* containing a common short tandem repeat (STR) polymorphism linked to autism (rs1880784044) resulted in an almost two-fold increase in midbrain enhancer activity (4/4 embryos;  $P=0.011$ ; Fig. 4e–g)<sup>68</sup>. These results indicate that dual-enSERT-2 can be used for rapid functional screening of non-coding variants linked to congenital disorders.

### Pathogenic enhancer variant activity at single-cell resolution

We next asked whether ectopic activity caused by gain-of-function enhancer variants can be quantitatively assigned to specific cell types in vivo using dual-enSERT. Such information, coupled with gene expression profiling can potentially reveal which genetic pathways lead to ectopic gene expression upon enhancer misregulation. We focused on the pathogenic hZRS<sup>404G>A</sup> variant allele for which the mechanism of ectopic *Shh* expression in the limb bud is not known (Fig. 1c)<sup>41</sup>. We performed two separate single-cell RNA-sequencing (scRNA-seq) experiments on dissected E11.5 hindlimb buds from F2 dual-enSERT-1 and F0 dual-enSERT-2 embryos, respectively (Figs. 1c, 3b, 5a and Supplementary Fig. 6a). In both transgenic embryos, hZRS<sup>5'ef</sup> allele drove *mCherry*, while hZRS<sup>404G>A</sup> allele drove *eGFP*. To decrease reporter gene dropout<sup>70,71</sup>, we adopted a nested PCR strategy to amplify *mCherry* and *eGFP* transcripts in our barcoded libraries (Methods and Fig. 5a)<sup>72</sup>. We processed scRNA-seq datasets generated from dual-enSERT-1 and dual-enSERT-2 hindlimb buds independently (Supplementary Fig. 6a). Supervised clustering of dual-enSERT-1 hindlimb produced thirteen distinct cell types, including a large mesenchymal cluster defined by the specific expression of well-known marker genes (Supplementary Fig. 6b and Supplementary Data File 3)<sup>73,74</sup>. We recovered every cell type in dual-enSERT-2 hindlimb, except for proximal cells which we excluded during hindlimb dissection for dual-enSERT-2 (Supplementary Fig. 6b). Having shown the reproducibility between dual-enSERT-1 and -2, we combined and integrated the two scRNA-seq datasets together to obtain ~21,000 cells and maximise statistical power for downstream analyses (Fig. 5b and Supplementary Fig. 6c).

We first asked in which cell types are the reference and variant hZRS alleles active based on reporter gene expression. The strongest cluster of *mCherry* expression, driven by the reference hZRS allele, was distal posterior mesenchyme, matching the expression of its target gene *Shh* (Fig. 5c–e). We also detected *mCherry* expression in immune cells where *Shh* is not expressed, possibly due to *hsp68* promoter activity<sup>20</sup>. By contrast, strong levels of *eGFP* expression driven by the hZRS<sup>404G>A</sup> allele were also detected in distal middle, anterior, and proximal anterior mesenchymal cells, matching the distribution of ectopic eGFP fluorescence in live embryos (Fig. 5c, d). These cell-type-specific expression patterns were consistent between F2 dual-enSERT-1 and F0 dual-enSERT-2 datasets (Supplementary Fig. 6d).

We next examined gene expression in cell subpopulations in which hZRS is normally active (*mCherry*+/*eGFP*+), ectopically active (*eGFP*+/*mCherry*-) and inactive (*mCherry*-/*eGFP*-). We performed unbiased differential gene expression analysis between these cell subpopulations to identify candidate genetic pathways linked to normal and ectopic *Shh* expression. With only thirty-six differentially expressed genes (FDR < 0.05, log<sub>2</sub>FC > ± 1.5), both normal and ectopic hZRS domains showed strong similarity in their transcriptional profiles, including enrichment of known mesenchymal transcription factors such as *Msx1*, *Lhx2*, *Lhx9*, *Twist1* and others (Fig. 5e). This is consistent with the fact that many transcription factors specifying mesenchymal fate are expressed in the entire “progress zone” beneath

the apical ectodermal ridge (AER) which includes ectopic anterior and normal posterior domains of hZRS activity<sup>75,76</sup>. By contrast, inactive cells expressed chondrocyte-specifying transcription factors like *Shox2* and *Sox9* (Fig. 5e)<sup>77,78</sup>. Differential gene expression analysis between three clusters identified only a few genes, such as *Asb4* that were specifically expressed in ectopic *eGFP*+/*mCherry*- cells.

To increase the sensitivity of the analysis, we subset cells expressing *mCherry* and/or *eGFP* and re-clustered them (Fig. 5f). We found that cells clustered into those expressing both *mCherry* and *eGFP* (normal domain, Clusters 1 and 2) and those only expressing *eGFP* (ectopic domain, Cluster 3) (Fig. 5f, g). To identify genes specifically upregulated in *eGFP*-expressing cells, we performed differential analysis between Cluster 3 and all other clusters. We identified several anterior-biased genes, including *Asb4*, *Pax9*, and *Hpse2* (Fig. 5h), which match their expression patterns by in situ hybridisation experiments of the limb bud. Taken together, these results implicate candidate pathways in ectopic hZRS activity and highlight how dual-enSERT enables the capture of variant allele-labelled cells at cellular resolution.

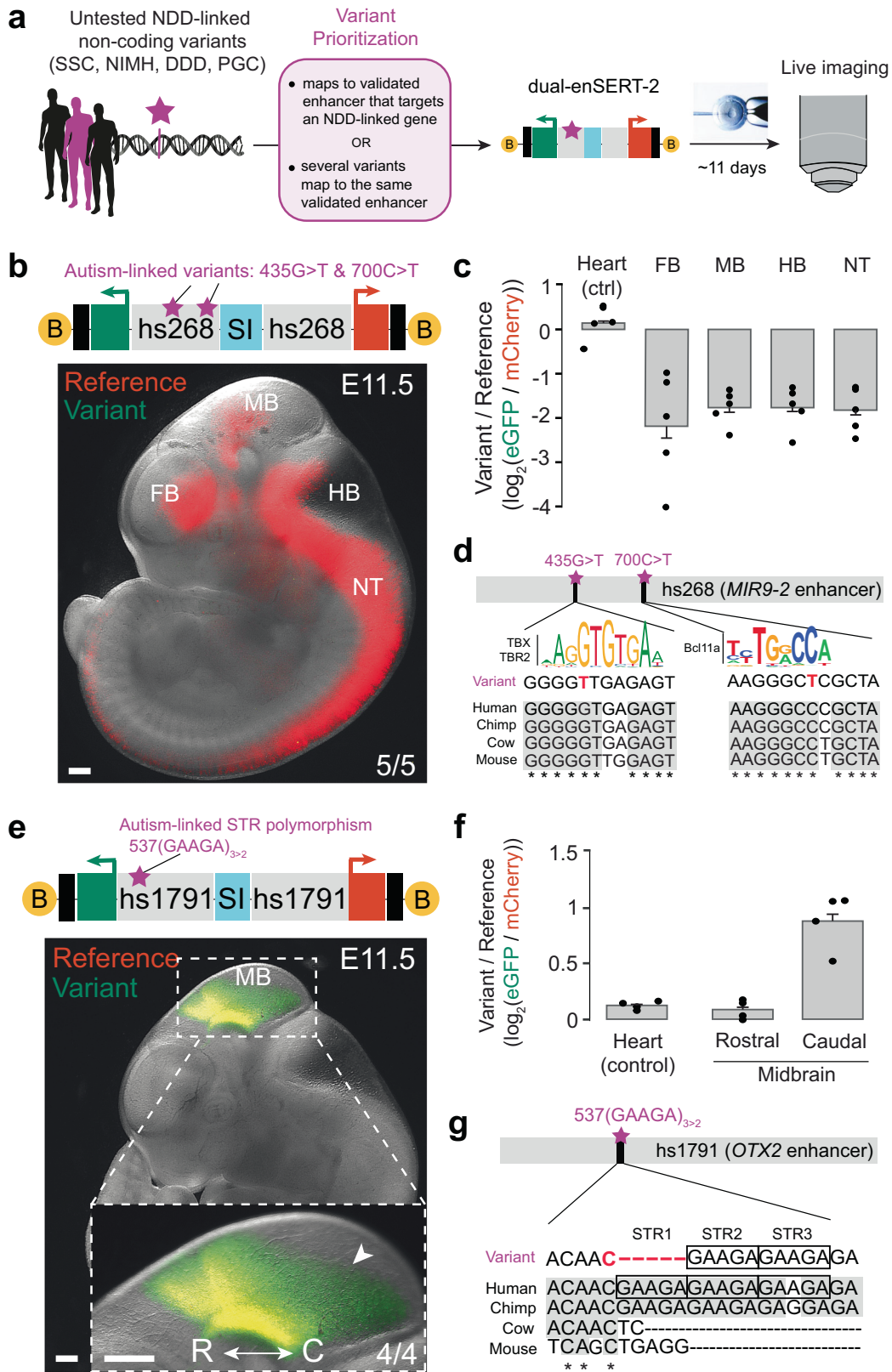
### Discussion

Realising the functional and therapeutic potential promised by large-scale human genomics studies depends on our ability to test the in vivo effects of candidate non-coding variants linked to human disease. However, no method currently exists for direct comparative testing of enhancer variant activity in a live mammal. In this study, we directly addressed this unmet need by developing dual-enSERT, a rapid, quantitative, and cost-effective method for simultaneous comparison of human enhancer allele activities in live mice. Dual-enSERT can be used to study rare and common non-coding variants and is compatible with a wide range of congenital disorders and organ systems. We also show that dual-enSERT can be easily combined with single-cell technologies to characterise gene expression in cells where an enhancer is ectopically active.

Our functional screening of previously uncharacterised candidate variants from WGS and GWAS shows reproducible effects on the in vivo enhancer activity of two out of seven tested enhancers. These effects were detected from just a few embryos by using quantifiable fluorophores and both enhancer alleles placed on one transgene, eliminating the impact of mosaicism seen in enSERT and traditional LacZ assays, which require large numbers of embryos to detect such effects<sup>17,20,79,80</sup>. The remaining five of seven variant enhancers did not change their activity in transgenic mice upon the introduction of these variants. It is possible that the variants that did not affect enhancer activity at E11.5 might have an impact during different developmental stages. Alternatively, these variants may be benign incidental observations in cases whose underlying disorder is caused by other genetic and/or environmental factors, which is often the case in human genetics studies<sup>20,81</sup>.

While developing dual-enSERT-2, we found that three copies of the widely used 5'-HS4 chicken β-globin insulator, each of which contains multiple CTCF sites<sup>24,57,58</sup>, unexpectedly failed to block communication between the hZRS enhancer and the *hsp68* minimal promoter (Supplementary Fig. 4b, c). These results support previous observations that insulator function can be locus- or enhancer-specific<sup>82–84</sup>. There is also evidence that insulator function is transcription factor-specific, i.e., some insulators work with one type of enhancer but not with another<sup>83,84</sup>. Therefore, caution should be taken when using individual insulators to protect transgenes in genome engineering applications.

Our synthetic insulator, created by fusing multiple copies of three of the most well-studied vertebrate insulators, A2, ALOXE3, and 5'-HS4, effectively blocked promoter interactions for five enhancers in the context of a transgene. We speculate the ability of SI to efficiently block enhancer-promoter interactions may derive from combining different mechanisms of insulation: A2 and 5'-HS4 depend on CTCF



while ALOXE3 relies on two B-boxes that recruit RNA polymerase III<sup>61,85,86</sup>. This SI has the potential to work in a wider number of genomic contexts and applications.

Beyond studying disease-linked enhancer variants, dual-enSERT can be used for other applications. Sequence divergence in enhancers is hypothesised to be a major driver of morphological and functional

evolution<sup>87,88</sup>; however, pinpointing and functionally testing the causal regulatory regions has been challenging. With dual-enSERT, the activity of candidate evolution-driving enhancers can be directly compared to a reference enhancer in the same animal to detect any functional changes. Dual-enSERT also provides a more time-effective and quantitative method for in vivo mutagenesis of enhancers that



**Fig. 4 | In vivo testing of uncharacterised variants from patients with autism spectrum disorder.** **a** Schematic depicting the identification and testing of uncharacterised human enhancer variants linked to neurodevelopmental disorders (NDDs). DDD, Deciphering Developmental Disorders; NIMH, National Institute of Mental Health cohort; PGC, Psychiatric Genomics Consortium; SSC, Simons Simplex Collection. **b** Fluorescent image of B-hs268<sup>ref</sup>-*mCherry*/SI/hs268<sup>var</sup>-*eGFP*-B whole embryo at E11.5. B, biotin. Scale bars, 500  $\mu$ m. **c** Plots quantifying fold-change ( $\log_2$ ) difference in reporter intensity between variant and reference alleles by tissue in B-hs268<sup>ref</sup>-*mCherry*/SI/hs268<sup>var</sup>-*eGFP*-B embryos. Data points represent independent biological replicates ( $n = 5$  embryos). Two-sided paired  $t$  tests vs. Heart: Forebrain (FB),  $P = 0.0011$ ; Midbrain (MB),  $P = 0.0093$ ; Hindbrain (HB),  $P = 0.0065$ ; Neural Tube (NT),  $P = 0.0056$ . Data represented as mean  $\pm$  SEM. **d** Putative TF motifs within the hs268 enhancer of *MIR9-2* that overlap with the loss-

of-function patient SNVs at the 435 and 700 nucleotide positions with conservation. **e** Fluorescent image of B-hs1791<sup>ref</sup>-*mCherry*/SI/hs1791<sup>var</sup>-*eGFP*-B whole embryo and high-resolution inset of the midbrain (MB) at E11.5. B, biotin; C, caudal; R, rostral. Scale bars, 500  $\mu$ m. **f** Plots quantifying fold-change ( $\log_2$ ) difference in reporter intensity between variant and reference alleles by tissue in B-hs1791<sup>ref</sup>-*mCherry*/SI/hs1791<sup>var</sup>-*eGFP*-B embryos. Data points represent independent biological replicates ( $n = 4$  embryos). Two-sided paired  $t$  tests vs. Heart: Rostral Midbrain,  $P = ns$ ; Caudal Midbrain,  $P = 0.011$ . Data represented as mean  $\pm$  SEM. **g** A common polymorphism variant at the 537-nucleotide position causes a reduction in short tandem repeat copy-number of GAAGA in the hs1791 enhancer of *OTX2*. Human silhouette and microscope cartoons reproduced courtesy of Zane Mitreva and Augustin Carpaneto, respectively (<http://sci-draw.io>). Source data are provided as a Source Data file.

**Table 1 | Summary of dual-enSERT-2 results for previously uncharacterised human enhancer variants**

Variant Location (hg38)	dbSNP ID	Associated Disorder	Enhancer (Putative Target Gene(s))	Normal Enhancer Activity	Reproducible Changes in Enhancer Activity	Reference
chr1:213425381T>C, chr1:213425533A>C	– –	Autism spectrum disorder	hs204 ( <i>PTPN14</i> )	Forebrain	No change	Shin et al. <sup>26</sup>
chr5:88396771G>T, chr5:88397035C>T	rs576375513, –	Autism spectrum disorder	hs268 ( <i>MIR9-2</i> )	Forebrain, Midbrain, Hindbrain, and Neural Tube	Loss of activity in the Forebrain, Midbrain, Hindbrain and Neural Tube	Shin et al. <sup>26</sup>
chr19:30843449G>C, chr19:30843509G>A	– –	Neurodevelopmental disorder	hs430 (-)	Midbrain	No change	Short et al. <sup>67</sup>
chr3:180462583T>C, chr3:180462704A>G	rs189450851, –	Neurodevelopmental disorder	hs655 ( <i>RNF220</i> , <i>ER13</i> , <i>DMAP1</i> )	Hindbrain	No change	Short et al. <sup>67</sup>
chr3:147847042T>C, chr3:147847133T>C, chr3:147847216T>C	–, –, rs1166255572	Autism spectrum disorder	hs1573 (-)	Forebrain, Hindbrain	No change	Shin et al. <sup>6</sup>
chr14:57007962CGAAGA>C	rs1880784044	Autism spectrum disorder	hs1791 ( <i>OTX2</i> )	Midbrain	Gain of activity in the Midbrain	Grove et al. <sup>68</sup>
chr15:66774248C>T, chr15:66774318C>T, chr15:66774321C>T	–, –, rs546228412	Autism spectrum disorder	hMM1518 ( <i>SMAD6</i> , <i>SCARLETLTR</i> )	Not Active	No change	Zhou et al. <sup>97</sup>

allows a whole embryo readout of changes in enhancer activity. Another potential application of dual-enSERT is the genetic labelling of specific cell populations. It is often difficult to isolate a desired cell population with a single genetic driver but using an intersectional strategy with two fluorescent reporters driven by different enhancers can enable labelling and isolation of smaller cellular populations<sup>89–91</sup>.

When testing non-mouse enhancers with dual-enSERT, we cannot rule out the potential effects caused by *trans*-regulatory divergence between different species, especially if no detectable changes in variant enhancer activity are observed<sup>20</sup>. In addition, dual-enSERT has a limited throughput which only allows testing two enhancer alleles per construct. Nevertheless, dual-enSERT provides a valuable addition to high-throughput methods such as MPRA, as changes in enhancer activities are detected in vivo in whole live mice and in a native chromosomal context. In summary, our work demonstrates the power of mouse transgenesis by enabling rapid and quantitative comparative in vivo testing of disease-linked variants for the interpretation of new human genetics findings.

## Methods

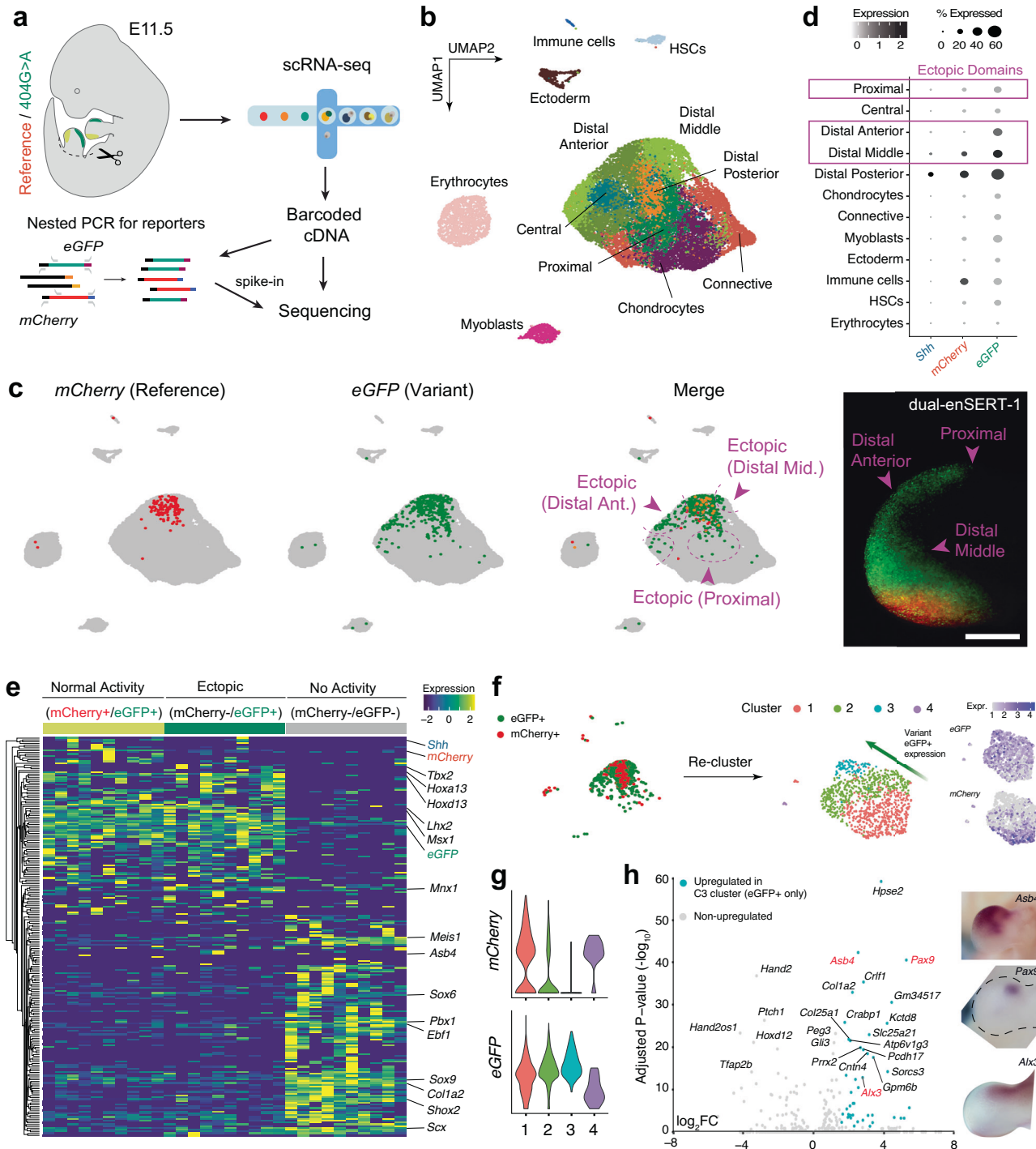
### Ethics statement

This research complies with all relevant ethical regulations. All animal procedures, including those related to the generation of transgenic animals, were conducted in accordance with the guidelines of the National Institutes of Health (NIH) and approved by the Institutional Animal Care and Use Committee at the University of California, Irvine under protocol no. AUP-23-005.

### Cloning of dual-enSERT constructs

**Dual-enSERT-1 plasmid construction.** To create dual-enSERT-1 constructs, we used PCR4-Hsp68::lacZ-H11 plasmid<sup>20</sup> (Addgene #139099). We replaced *lacZ* with *eGFP* or *mCherry* fluorescent reporters using Gibson cloning<sup>92</sup>. The resulting constructs contained *eGFP* (PCR4-Hsp68::eGFP-H11) or *mCherry* (PCR4-Hsp68::mCherry-H11) fluorescent reporter genes under the control of the *hsp68* minimal promoter and homology arms targeting the H11 locus. Each tested enhancer sequence was cloned into the corresponding dual-enSERT-1 vector using NotI digestion followed by Gibson assembly (NEB, E2611)<sup>38,92</sup>. Reference allele sequences for hZRS and hs737 enhancers were obtained via PCR cloning from human genomic DNA (Promega, G304A). Primers used for each enhancer sequence are outlined in Supplementary Data File 2. All PCR cloning was performed using Q5 High-Fidelity Polymerase (NEB, M0491) or KOD polymerase (Toyobo, #KMM-201). Variant allele sequences for hZRS<sup>404G>A</sup>, hZRS<sup>446T>A</sup>, and hs737<sup>830G>A</sup> were synthesised as gBlocks (Integrated DNA Technologies (IDT)).

**Dual-enSERT-2 plasmid construction.** The bicistronic plasmid was designed as a modified version of the fluorescent dual-enSERT-1 plasmid. The cHS4 insulator sequence was cloned from genomic chicken DNA (Zyagen, GC-314) using the primers from Bhatia et al. To obtain three copies, linker sequences (fragments of the LacZ gene) were added as overhangs to the primers. The A2 insulator was synthesised as a gBlock using the sequence reported by Liu et al.<sup>61</sup> while ALOXE3 was cloned from human genomic DNA using primers reported by Raab



**Fig. 5 | Characterisation of pathogenic enhancer variant activity at single-cell resolution.** **a** Single-cell transcriptomic profiling of E11.5 hindlimbs from transgenic mice in which hZRS<sup>ref</sup> drives *mCherry* and hZRS<sup>404G>A</sup> drives *eGFP*. A nested PCR strategy was used to amplify *mCherry* and *eGFP* reporter transcripts (Methods). **b** Integrated UMAP plot showing ~21,000 cells clustered by cell type. Mesenchymal clusters are defined by the expression of spatially-defined genes. HSCs, hematopoietic stem cells. **c** Feature plots showing *mCherry* and *eGFP* expression with overlapping cells marked in yellow. Areas of ectopic *eGFP* expression are highlighted in magenta with an accompanying fluorescent image on the right. Data derived from one dual-enSERT-1 and one dual-enSERT-2 replicate. Scale bar, 500  $\mu$ m. **d** Dot plot quantifying percentage and normalised expression of cells expressing *eGFP*, *mCherry*, and *Shh* within each cluster. Clusters with ectopic *eGFP*

expression are highlighted. **e** Heatmap of differentially expressed genes between *mCherry* + (normal hZRS activity), *mCherry*-*eGFP* + (ectopic hZRS activity), and *mCherry*-*eGFP*- (inactive ZRS) cell subpopulations. Unsupervised hierarchical clustering of genes on the left; select marker genes on the right. **f** UMAP plots of reclustered *mCherry* + and *eGFP* + cells with accompanying feature plots depicting their expression. **g** Violin plots quantifying *mCherry* and *eGFP* expression across clusters. **h** Volcano plot depicts differential gene expression between Cluster 3 and the other clusters. Genes upregulated (Adjusted *P*-value < 0.05 and log<sub>2</sub>FC > 2) in Cluster 3 are coloured in teal. Non-parametric Wilcoxon-rank sums test. The spatial distribution of representative marker genes in E11.5 limb buds is shown on the right. Images reproduced with permission from the Embryos database (<http://embryos.jp>).

et al.<sup>60</sup>. Fusion PCR was performed to obtain the final synthetic insulator fragment consisting of two copies of A2, one copy of ALOXE3, and two copies of cHS4. Enhancer-*hsp68p*-reporter sequences were then PCR amplified from dual-enSERT-1 plasmids.

To streamline the cloning of different enhancers into dual-enSERT-2 plasmids, NotI and AgeI restriction digestion sites were added to the outside of the two enhancer sites via PCR. Dual-enSERT-2 plasmid was digested with NotI (NEB, R3189) and AgeI (NEB, R3552) to create an empty vector without enhancers. Then, plasmids were assembled using a four-fragment Gibson-based assembly of (i) empty dual-enSERT-2 backbone, (ii) reference enhancer allele, (iii) variant reference allele, and (iv) synthetic insulator. The hZRS and hs737 reference and variant allele sequences were PCR amplified from dual-enSERT-1 plasmids. All other reference enhancer alleles were cloned from human genomic DNA using primers listed in Supplementary Data File 2. When sequence complexity was sufficiently low for in vitro synthesis, variant enhancer alleles were synthesised as gBlocks (IDT). If unable to be synthesised, custom primers with overhangs containing the selected variants were designed and used to clone separate fragments of the enhancer. Gibson-based cloning methods were then used to assemble the full-length variant enhancer alleles.

To enable linearisation of the dual-enSERT-2 donor plasmid, Paul (also known as BssHII) sites were added to the outside of the H11 homology arms via PCR. A dual-enSERT-2 plasmid (2 µg) was digested overnight with Paul (NEB, R0199) at 50 °C in rCutSmart Buffer. The following day, the reaction was inactivated by incubation at 65 °C for 15 min. To end-fill the 3' overhangs with biotinylated nucleotides, Biotin-11-GTP (100 µM; Jena Bioscience, NU-971-BIOX) and biotin-11-CTP (100 µM; Jena Bioscience, NU-831-BIOX) were added to the reaction with T4 DNA Polymerase (1 unit; NEB, M0203) and incubated at 12 °C for 15 min. The reaction was stopped by the addition of EDTA (100 mM), and heat inactivated at 75 °C for 20 min. Biotinylation of fragments was confirmed by pull-down with streptavidin T1 Dynabeads (Thermo Fisher Scientific, cat. no. 65601).

For all constructs in this study, restriction digestion with SacI, Eco72I, NotI and/or AgeI (Thermo Fisher, FastDigest), Sanger and/or whole-plasmid (Plasmidsaurus) sequencing were performed to ensure the integrity of the vector and enhancer sequences before zygote microinjection. See Supplementary Table 1 and Supplementary Fig. 7 for complete details of all plasmids created and used in this study, of which the construct backbones are available at Addgene (#211940, #211941 and #211942).

### Assessment of 5'-HS4 insulator for dual-enSERT-2

We first tested whether three copies of the previously characterised chicken  $\beta$ -globin insulator, 5'-HS4, could prevent the cross-activation of two enhancer-reporter transgenes. 5'-HS4 is widely used for its robust ability to block enhancer-promoter activation in the genome<sup>57–59</sup> and in the context of a zebrafish transgene<sup>24</sup>. We additionally placed three copies of the 5'-HS4 insulator (3xHS4) into the plasmid backbone to prevent cross-activation between different copies of the transgene in the event of multi-copy integrations at the H11 landing site<sup>20</sup> (Supplementary Fig. 4a). We injected the resulting hZRS<sup>ref</sup>-*mCherry*/3xHS4/hZRS<sup>404G>A</sup>-*eGFP* bicistronic construct into mouse zygotes and collected transgenic embryos at E11.5. We detected robust mCherry and eGFP expression in anterior cells in all examined transgenic embryos, indicating that the variant hZRS allele can simultaneously activate *eGFP* and *mCherry* reporter genes in this transgene context (5/5 of embryos with a single-copy transgene integration at H11; 11/11 of embryos with multi-copy transgene integration at H11, Supplementary Fig. 4b, c and Supplementary Data File 1). Quantitatively, we found no difference in fluorescent intensity across the anterior and ZPA regions of the fore- and hindlimb (Forelimb ZPA,  $P = ns$ ; Forelimb Anterior,  $P = ns$ ; Hindlimb ZPA,  $P = ns$ ; Hindlimb

Anterior,  $P = ns$ ; Supplementary Fig. 4b, c). These results show that three copies of the 5'-HS4 insulator are insufficient to insulate the variant hZRS allele from cross-activating the other reporter gene (Supplementary Fig. 4j).

### Transgenic mouse generation

All transgenic mice in this study were generated using a CRISPR/Cas9 microinjection protocol, as previously described<sup>20,93</sup>. Briefly, a mix of (i) Cas9 protein (final concentration of 20 ng/µl; IDT Cat. No. 1074181), (ii) sgRNA (50 ng/µl) and (iii) circular donor plasmid (7 ng/µl) or linearised fragment (1 ng/µl) in injection buffer (10 mM Tris, pH 7.5; 0.1 mM EDTA) was injected into the pronucleus of FVB embryos. All donor plasmids or fragments were column purified using a PCR purification kit (Qiagen) and eluted into an injection buffer before injection. Female mice (CD-1 strain) were used as surrogate mothers. Superovulated female FVB mice (7–8 weeks old) were mated to FVB stud males, and fertilised embryos were collected from oviducts. The injected zygotes were cultured in M16 with amino acids at 37 °C under 5% CO<sub>2</sub> for approximately 2 hr. Afterwards, zygotes were transferred into the uterus of pseudopregnant CD-1 females. F0 embryos were either brought to gestation (dual-enSERT-1) or collected at E11.5 (dual-enSERT-2).

### Mouse strains and embryo collection

Mice were maintained in standard housing conditions (temperature between 19–23 °C and humidity between 40–60%) on a reversed 12 h dark–light cycle with food and water provided *ad libitum*. Time of gestation was identified by the presence of vaginal sperm plugs, indicating E0.5. Pregnant dams were humanely euthanized, and E11.5 embryos were carefully removed under brightfield stereoscopes in ice-cold PBS (Cytiva, SH30256.01). Both sexes of embryos were presumed to be included. Yolk sacs or tail pieces were collected for genotyping. Successful integration events at the H11 locus were determined by PCR using primers described previously<sup>20,38</sup>.

### Live fluorescent imaging and quantification

Embryos were imaged in ice-cold PBS in a small petri dish (Greiner Bio-One, #627102) atop a thin layer of 2% gel agarose (Fisher, BP160). Images were taken on a Zeiss V8 stereoscope using a monochromic camera (AxioCam 202, Zeiss), fibre optic light source (Zeiss, CL1500) LED fluorescent laser (X-Cite, Xylis), and fluorescent channels at 488 and 555 nm wavelengths. Single-channel images were merged using Zeiss BioLite software. Quantification of fluorescent reporter intensity was performed by importing the original.czi files into Fiji software<sup>94</sup>. Regions of interest were outlined on the variant-driven colour channel and then transferred to the reference-driven colour channel to measure the mean fluorescence intensities for each region. To account for intrinsic differences in fluorophore intensity and background, the mean fluorescence of embryo tissues with no enhancer activity was measured, averaged, and subtracted from the regions-of-interest fluorescent intensities. Because the *hsp68* promoter causes leaky reporter activity in the heart<sup>20</sup>, the heart was used as a negative control in two-sided, paired *t* tests to account for differences in fluorophore maturation time and half-life.

### Fluorescence-activated cell sorting

After imaging, the anterior portions of dual-enSERT-1 mouse limb buds carrying two hZRS reporter transgenes were carefully dissected under the fluorescent scope. Dissected regions from each embryo were pooled separately and then incubated with collagenase II (Gibco, #17101015, 0.2 µL at 100 u/µl) for 10 min at 700 rpm and 37 °C with trituration every 5 min with a P200 pipette. 450 µL of 10% FBS (Thermo Fisher, #A3840201) was added and dissociated cells were spun down at 500 × *g* for 5 min. Cells were resuspended in 200 µL of 0.04% BSA (Millipore Sigma, #A1595) and filtered using 40 µm P1000 Flowmi cell

filters (SP Bel-Art, #136800040). After gating with forebrain tissue as a negative control, mCherry<sup>+</sup>, eGFP<sup>+</sup>, and double-positive cell populations were quantified using a FACSria Fusion Sorter (BD Biosciences). Fisher exact tests were performed between genotypes of double-positive, mCherry<sup>+</sup>, and eGFP<sup>+</sup> cells.

### RNA isolation and cDNA preparation

After fluorescent-activated cell sorting, 10  $\mu$ L of RNeasy Protect (Qiagen, #76104) was immediately added to cells. RNA was isolated using the RNeasy Mini Kit (Qiagen, #74104) according to manufacturer instructions. cDNA libraries were constructed from isolated RNA with the ProtoScript<sup>®</sup> II First Strand cDNA Synthesis Kit (NEB, #E6560) using Oligo(d)T primers and following the manufacturer's recommended protocol.

### Quantitative PCR

Quantitative PCR was performed for *mCherry* and *eGFP* transgenes to quantify mRNA expression (from cDNA) and determine the copy-number of enhancer-reporter constructs in each dual-enSERT-1 mouse line (from genomic DNA). From mouse genomic DNA, *mCherry* or *eGFP* transgenes and endogenous control of known copy-number were amplified by primers and fluorescent probes (FAM or HEX, IDT), designed using PrimerQuest Tool (IDT) (Supplementary Data File 2). Reactions were performed in multiplex and carried out with Prime-Time Gene Expression Master Mix (IDT, 105577) in a C1000 Touch Thermal Cycler with a CFX96 Real-Time System module (Bio-Rad, 1845096). Cycle threshold values (Ct) for each amplicon were extracted from .zpcr files using CFX Maestro Software (Bio-Rad). Transgene copy numbers (normalised to endogenous control) were calculated using a modified  $2^{-\Delta\Delta CT}$  method in which samples of unknown copy numbers were compared to positive control samples containing verified single transgene insertions. To quantify absolute mRNA expression from cDNA, *mCherry* or *eGFP* were normalised to the housekeeping gene *Gapdh* using the  $2^{\Delta CT}$  method<sup>95</sup>. Plots were generated using GraphPad Prism, version 10.1.2.

### Prioritisation of disease-linked enhancer variants for dual-enSERT-2

To select previously uncharacterised human enhancer variants for testing with dual-enSERT-2, we extracted published rare and common variants identified from recent GWAS and WGS studies on patients with neurodevelopmental disorders<sup>17,26,65–68</sup>. The genomic locations of these variants were then intersected with the coordinates of mouse and human enhancers experimentally validated in vivo<sup>36,37</sup>. Enhancer variants were prioritised by either multiple independent variants mapping to the same enhancer or the enhancer interacting with the promoter of developmental disorder-lined gene based on published capture Hi-C data<sup>69</sup>.

### Single-cell transcriptomics

Hindlimb buds from an hZRS<sup>ref</sup>-*mCherry*/hZRS<sup>404G>A</sup>-*eGFP* E11.5 (dual-enSERT-1) and B-hZRS<sup>ref</sup>-*mCherry*/SI/hZRS<sup>404G>A</sup>-*eGFP*-B (dual-enSERT-2) mouse embryo were dissected in ice-cold PBS and processed independently. A single-cell suspension was obtained using collagenase II, as described above for FACS. Dissociated cells were resuspended in 25  $\mu$ L of 0.04% BSA before being quantified and inspected for viability using Trypan Blue (Bio-Rad, #1450013). Live cells were counted by hemocytometer (Bio-Rad, #1450011) and loaded at a concentration that would enable recovery of 10,000 nuclei by the Chromium Next GEM Chip G Single Cell Kit 3' Gene Expression Kit v3.1 (10X Genomics, cat no. 1000127). Captured cells were pair-end sequenced on an Illumina NovaSeq 6000 for ~50,000 reads per cell.

To amplify reporter transcripts, nested PCR for *mCherry* and *eGFP* transcripts was performed on the prepared cDNA library as described previously, with slight modifications<sup>72</sup>. The first PCR included a trimer

mix of *mCherry*- and *eGFP*-specific forward primers (mCherry-1; eGFP-1) and an R1-targeting reverse primer (Supplementary Data File 2). After bead clean-up (CleanNA, CNGS-0050), a second PCR reaction was performed on the PCR product using a trimer mix targeting *eGFP* and *mCherry* (mCherry-2; eGFP-2) and the same R1 reverse primer (Supplementary Data File 2). The resulting DNA was then bead-purified in preparation for the final PCR using the i5:i7 indices from Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 (10X Genomics, cat. no 1000128). All PCR reactions were performed using Q5 polymerase (NEB, M0491). The final indexed and purified DNA was spiked in for additional sequencing of 5000 reads per cell.

Fastq files were aligned to a modified mm39 genome assembly (Ensembl, GCA\_000001635.9) that included *mCherry* and *eGFP* sequences and barcodes were counted to generate a count matrix using Cell Ranger (10x Genomics, Cell Ranger 3.1.0). Count matrix data were analysed using the Seurat R package, version 4<sup>96</sup>. A SeuratObject was created, and quality control was performed to exclude cells with greater than 5% mitochondrial DNA and with UMI counts between 2500 and 8000 per cell. Transcriptome data were normalised and scaled, and variably expressed genes ( $n=2000$ ) were utilised for principal component analysis. An elbow plot was produced to calculate the number of dimensions. Because the developing E11.5 limb bud is highly proliferative, cell cycle genes (S and G2M genes from Seurat) were regressed to enhance the detection of cell clusters based on spatial gene expression. Nearest neighbour, unsupervised clustering and UMAP analysis were then performed (dimensions=13, resolution=0.5). Cell types were identified from the resulting clusters using well-defined marker genes<sup>73,74</sup>. FeaturePlot was utilised to produce the overlaid plots of *mCherry* and *eGFP* expression. The DotPlot function was used to generate expression data by each cell cluster. To determine the topmost upregulated genes within the normal and ectopic hZRS domains, cells were selected for expression of the reporter transcripts: either *mCherry*>2 (normal hZRS) or *eGFP*>2 (variant hZRS), or inactive (remaining cells). Differential gene expression was defined by comparing cell populations using the FindMarkers function, and the heatmap was downsampled ( $n=100$ ) for easier visualisation. All barcode meta-data, including cell annotations and version of dual-enSERT, are reported in Supplementary Data File 3.

### Statistics and reproducibility

Statistical analyses were performed using R version 4.3.1 and Microsoft Excel version 16.79.1. Experimental parameters, including the number of embryos, statistical tests, and significance are reported in the text, figures and/or figure legends. The investigators were blinded to genotype for all imaging and quantification analyses. No statistical method was used to predetermine sample size, no data were excluded from the analyses, and no randomisation was performed. *P*-values less than 0.05 were considered significant. All bar graphs are shown as mean  $\pm$  SEM. Raw sequencing data were analysed on the UCI high-performance computing cluster.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Processed and raw scRNA-seq data in this study have been deposited in the GEO database under accession code: [GSE244244](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE244244). Source data are provided in this paper.

### References

- Westra, H.-J. & Franke, L. From genome to function by studying eQTLs. *Biochim. Biophys. Acta* **1842**, 1896–1902 (2014).
- Claussnitzer, M. et al. A brief history of human disease genetics. *Nature* **577**, 179–189 (2020).

3. Lappalainen, T. & MacArthur, D. G. From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021).
4. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**, 779–797 (2013).
5. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
6. Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
7. Turro, E. et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96–102 (2020).
8. Dong, S. et al. Annotating and prioritizing human non-coding variants with RegulomeDB v.2. *Nat. Genet.* **55**, 724–726 (2023).
9. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
10. Gazal, S. et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet.* **50**, 1600–1607 (2018).
11. Ulirsch, J. C. et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019).
12. Claringbould, A. & Zaugg, J. B. Enhancers in disease: molecular basis and emerging treatment strategies. *Trends Mol. Med.* **27**, 1060–1073 (2021).
13. Pachano, T., Haro, E. & Rada-Iglesias, A. Enhancer-gene specificity in development and disease. *Development* **149**, <https://doi.org/10.1242/dev.186536> (2022).
14. Lewis, A. et al. A polymorphic enhancer near GREM1 influences bowel cancer risk through differential CDX2 and TCF7L2 binding. *Cell Rep.* **8**, 983–990 (2014).
15. Clausnitzer, M., Hui, C.-C. & Kellis, M. FTO obesity variant and adipocyte browning in humans. *N. Engl. J. Med.* **374**, 192–193 (2016).
16. Doan, R. N. et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354 (2016).
17. Turner, T. N. et al. Genomic patterns of DE Novo mutation in simplex autism. *Cell* **171**, 710–722 (2017).
18. Putlyayeva, L. V. et al. The minor variant of the single-nucleotide polymorphism rs3753381 affects the activity of a SLAMF1 enhancer. *Acta Naturae* **9**, 94–102 (2017).
19. Eufrásio, A. et al. In vivo reporter assays uncover changes in enhancer activity caused by type 2 diabetes-associated single nucleotide polymorphisms. *Diabetes* **69**, 2794–2805 (2020).
20. Kvon, E. Z. et al. Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. *Cell* **180**, 1262–1271 (2020).
21. Yanchus, C. et al. A noncoding single-nucleotide polymorphism at 8q24 drives IDH1-mutant glioma formation. *Science* **378**, 68–78 (2022).
22. Spieler, D. et al. Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. *Genome Res.* **24**, 592–603 (2014).
23. Bhatia, S. et al. Functional assessment of disease-associated regulatory variants in vivo using a versatile dual colour transgenesis strategy in zebrafish. *PLoS Genet.* **11**, e1005193 (2015).
24. Bhatia, S. et al. Quantitative spatial and temporal assessment of regulatory element activity in zebrafish. *Elife* **10**, <https://doi.org/10.7554/elife.65601> (2021).
25. Bengani, H. et al. Identification and functional modelling of plausibly causative cis-regulatory variants in a highly-selected cohort with X-linked intellectual disability. *PLoS ONE* **16**, e0256181 (2021).
26. Shin, T. et al. Rare variation in non-coding regions with evolutionary signatures contributes to autism spectrum disorder risk. *Cell Genom* **4**, 100609 (2024).
27. Maricque, B. B., Chaudhari, H. G. & Cohen, B. A. A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. *Nat. Biotechnol.* **37**, 90–95 (2018).
28. Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**, 159–164 (2015).
29. Patwardhan, R. P. et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
30. White, M. A., Myers, C. A., Corbo, J. C. & Cohen, B. A. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. USA* **110**, 11952–11957 (2013).
31. Brown, A. R. et al. An in vivo massively parallel platform for deciphering tissue-specific regulatory function. Preprint at <https://doi.org/10.1101/2022.11.23.517755> (2022).
32. Lagunas, T. Jr et al. A Cre-dependent massively parallel reporter assay allows for cell-type specific assessment of the functional effects of non-coding elements in vivo. *Commun. Biol.* **6**, 1151 (2023).
33. Deng, C. et al. Massively parallel characterization of regulatory elements in the developing human cortex. *Science* **384**, eadh0559 (2024).
34. Capauto, D. et al. Characterization of enhancer activity in early human neurodevelopment using Massively Parallel Reporter Assay (MPRA) and forebrain organoids. *Sci. Rep.* **14**, 3936 (2024).
35. Kothary, R. et al. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707–714 (1989).
36. Pennacchio, L. A. et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
37. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
38. Osterwalder, M. et al. Characterization of mammalian in vivo enhancers using mouse transgenesis and CRISPR genome editing. *Methods Mol. Biol.* **2403**, 147–186 (2022).
39. Zakany, J., Tuggle, C. K., Patel, M. D. & Nguyen-Huu, M. C. Spatial regulation of homeobox gene fusions in the embryonic central nervous system of transgenic mice. *Neuron* **1**, 679–691 (1988).
40. Kvon, E. Z. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* **106**, 185–192 (2015).
41. Lettice, L. A., Hill, A. E., Devenney, P. S. & Hill, R. E. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* **17**, 978–985 (2008).
42. Fakhouri, W. D. et al. An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. *Hum. Mol. Genet.* **23**, 2711–2720 (2014).
43. Tasic, B. et al. Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc. Natl. Acad. Sci. USA* **108**, 7902–7907 (2011).
44. Snetkova, V. et al. Ultraconserved enhancer function does not require perfect sequence conservation. *Nat. Genet.* **53**, 521–528 (2021).
45. Hippenmeyer, S. et al. Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. *Neuron* **68**, 695–709 (2010).
46. Heutink, P. et al. The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. *Nat. Genet.* **6**, 287–292 (1994).
47. Lettice, L. A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
48. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).

49. Padhi, E. M. et al. Coding and noncoding variants in EBF3 are involved in HADDs and simplex autism. *Hum. Genom.* **15**, 44 (2021).
50. Pomerantz, M. M. et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.* **41**, 882–884 (2009).
51. Zhang, X., Cowper-Salari, R., Bailey, S. D., Moore, J. H. & Lupien, M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res.* **22**, 1437–1446 (2012).
52. Chatterjee, S. et al. Enhancer variants synergistically drive dysfunction of a gene regulatory network in Hirschsprung disease. *Cell* **167**, 355–368 (2016).
53. Dunning, A. M. et al. Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nat. Genet.* **48**, 374–386 (2016).
54. Khmelinskii, A. et al. Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers. *Mol. Biol. Cell* **27**, 360–370 (2016).
55. Balleza, E., Kim, J. M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods* **15**, 47–51 (2018).
56. Xu, C. et al. A novel ZRS variant causes preaxial polydactyly type I by increased sonic hedgehog expression in the developing limb bud. *Genet. Med.* **22**, 189–198 (2020).
57. Bungert, J. et al. Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev.* **9**, 3083–3096 (1995).
58. Yusufzai, T. M. & Felsenfeld, G. The 5′-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc. Natl. Acad. Sci. USA* **101**, 8620–8624 (2004).
59. Huang, H. et al. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat. Genet.* **53**, 1064–1074 (2021).
60. Raab, J. R. et al. Human tRNA genes function as chromatin insulators. *EMBO J.* **31**, 330–350 (2012).
61. Liu, M. et al. Genomic discovery of potent chromatin insulators for human gene therapy. *Nat. Biotechnol.* **33**, 198–203 (2015).
62. Bower, G. et al. Conserved Cis-acting range Extender element mediates extreme long-range enhancer activity in mammals. Preprint at <https://doi.org/10.1101/2024.05.26.595809> (2024).
63. Gutierrez-Triana, J. A. et al. Efficient single-copy HDR by 5′ modified long dsDNA donors. *Elife* **7**, <https://doi.org/10.7554/elife.39468> (2018).
64. Medert, R. et al. Efficient single copy integration via homology-directed repair (scHDR) by 5′ modification of large DNA donor fragments in mice. *Nucleic Acids Res.* **51**, e14 (2023).
65. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
66. Niemi, M. E. K. et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* **562**, 268–271 (2018).
67. Short, P. J. et al. De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611–616 (2018).
68. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
69. Chen, Z. et al. Increased enhancer-promoter interactions during developmental enhancer activation in mammals. *Nat. Genet.* **56**, 675–685 (2024).
70. Qiu, P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* **11**, 1169 (2020).
71. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742 (2014).
72. Pollina, E. A. et al. A NPAS4-NuA4 complex couples synaptic activity to DNA repair. *Nature* **614**, 732–741 (2023).
73. Desanlis, I., Paul, R. & Kmita, M. Transcriptional trajectories in mouse limb buds reveal the transition from anterior-posterior to proximal-distal patterning at early limb bud stage. *J. Dev. Biol.* **8**, 31 (2020).
74. Yokoyama, S. et al. Analysis of transcription factors expressed at the anterior mouse limb bud. *PLoS ONE* **12**, e0175673 (2017).
75. Towers, M. & Tickle, C. Growing models of vertebrate limb development. *Development* **136**, 179–190 (2009).
76. Markman, S. et al. A single-cell census of mouse limb development identifies complex spatiotemporal dynamics of skeleton formation. *Dev. Cell* **58**, 565–581 (2023).
77. Akiyama, H., Chaboissier, M.-C., Martin, J. F., Schedl, A. & Crombrugge, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* **16**, 2813–2828 (2002).
78. Yu, L. et al. Shox2 is required for chondrocyte proliferation and maturation in proximal limb skeleton. *Dev. Biol.* **306**, 549–559 (2007).
79. Tenney, A. P. et al. Noncoding variants alter GATA2 expression in rhombomere 4 motor neurons and cause dominant hereditary congenital facial paresis. *Nat. Genet.* **55**, 1149–1163 (2023).
80. Kosicki, M. et al. Massively parallel reporter assays and mouse transgenic assays provide complementary information about neuronal enhancer activity. Preprint at <https://doi.org/10.1101/2024.04.22.590634> (2024).
81. Gaulton, K. J., Preissl, S. & Ren, B. Interpreting non-coding disease-associated human variants using single-cell epigenomics. *Nat. Rev. Genet.* **24**, 516–534 (2023).
82. Uchida, N., Washington, K. N., Lap, C. J., Hsieh, M. M. & Tisdale, J. F. Chicken HS4 insulators have minimal barrier function among progeny of human hematopoietic cells transduced with an HIV1-based lentiviral vector. *Mol. Ther.* **19**, 133–139 (2011).
83. Ribeiro-Dos-Santos, A. M., Hogan, M. S., Luther, R. D., Brosh, R. & Maurano, M. T. Genomic context sensitivity of insulator function. *Genome Res.* **32**, 425–436 (2022).
84. Hong, C. K. Y. et al. Massively parallel characterization of insulator activity across the genome. *Nat. Commun.* **15**, 8350 (2024).
85. Farrell, C. M., West, A. G. & Felsenfeld, G. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol. Cell Biol.* **22**, 3820–3831 (2002).
86. Schramm, L. & Hernandez, N. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593–2620 (2002).
87. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: Transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
88. Rebeiz, M. & Tsiantis, M. Enhancer evolution and the origins of morphological novelty. *Curr. Opin. Genet. Dev.* **45**, 115–123 (2017).
89. He, M. et al. Strategies and tools for combinatorial targeting of GABAergic neurons in mouse cerebral cortex. *Neuron* **91**, 1228–1243 (2016).
90. Daigle, T. L. et al. A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* **174**, 465–480 (2018).
91. Matho, K. S. et al. Genetic dissection of the glutamatergic neuron system in cerebral cortex. *Nature* **598**, 182–187 (2021).
92. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
93. Kvon, E. Z. et al. Progressive loss of function in a limb enhancer during snake evolution. *Cell* **167**, 633–642.e11 (2016).
94. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

95. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
96. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
97. Zhou, J. et al. Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nat. Genet.* **51**, 973–980 (2019).

## Acknowledgements

The authors would like to acknowledge the UCI Transgenic Mouse Facility for help in the generation of transgenic mouse lines, Elizabeth Pollina (Washington University at St. Louis) and Daniel Gillam (Harvard) for their technical assistance with the nested PCR protocol, and Zane Mitrevica and Augustin Carpaneto (via sci-draw.io) for open-access use of human silhouette and microscope cartoons, respectively. This work was made possible, in part, through access to the Genomics Research and Technology Hub Shared Resource of the Cancer Centre Support Grant (P30CA-062203) at the University of California, Irvine and NIH shared instrumentation grants 1S1ORR025496-01, 1S10OD010794-01, and 1S10OD021718-01. The authors also thank Sabbi Lall (Life Science Editors) and Kvon lab members for their comments and suggestions on the manuscript. This work was supported by a National Institutes of Health grant R01HD115268 (to E.Z.K.), F30HD110233 (E.W.H.), T32NS082174 (E.W.H.), and T32GM008620 (E.W.H.) from the National Institutes of Health.

## Author contributions

E.W.H. and E.Z.K. conceived the project. E.W.H., T.A.L., S.H.J. and C.X.C. performed mouse transgenesis experiments, imaging and analysed the data. E.W.H. performed single-cell RNA-seq experiments and analysed the data. J.A.A. performed qPCR experiments and analysed the data. E.W.H. and E.Z.K. wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-55500-7>.

**Correspondence** and requests for materials should be addressed to Evgeny Z. Kvon.

**Peer review information** *Nature Communications* thanks Sumantra Chatterjee, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025