Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning

Steven A. Vokes[1], Hongkai Ji[2,3], Scott McCuine[4], Toyoaki Tenzen[1], Shane Giles[4], Sheng Zhong[3,*], William J. R. Longabaugh[5], Eric H. Davidson[6], Wing H. Wong[3] and Andrew P. McMahon[1,7,†]

Sonic hedgehog (Shh) acts as a morphogen to mediate the specification of distinct cell identities in the ventral neural tube through a Gli-mediated (Gli1-3) transcriptional network. Identifying Gli targets in a systematic fashion is central to the understanding of the action of Shh. We examined this issue in differentiating neural progenitors in mouse. An epitope-tagged Gli-activator protein was used to directly isolate cis-regulatory sequences by chromatin immunoprecipitation (ChIP). ChIP products were then used to screen custom genomic tiling arrays of putative Hedgehog (Hh) targets predicted from transcriptional profiling studies, surveying 50-150 kb of non-transcribed sequence for each candidate. In addition to identifying expected Gli-target sites, the data predicted a number of unreported direct targets of Shh action. Transgenic analysis of binding regions in Nkx2.2, Nkx2.1 (Titf1) and Rab34 established these as direct Hh targets. These data also facilitated the generation of an algorithm that improved in silico predictions of Hh target genes. Together, these approaches provide significant new insights into both tissue-specific and general transcriptional targets in a crucial Shh-mediated patterning process.

KEY WORDS: Chromatin immunoprecipitation, Gli, Neural patterning, Sonic hedgehog, Mouse

INTRODUCTION

The Hedgehog (Hh) signaling pathway represents one of approximately six core signaling pathways that dictate development in most bilateria (Davidson and Erwin, 2006; Hooper and Scott, 2005). Hh ligands regulate diverse processes, including morphogen-mediated patterning, cell cycle regulation and cell polarity, operating in a variety of cellular contexts from the forming segments of the early *Drosophila* embryo to the digit-forming mesenchyme of the vertebrate limb (McMahon et al., 2003). Genetic and biochemical studies indicate that all Hh signaling is likely to be mediated by Ci/Gli zinc finger domain-containing transcription regulators, corresponding in vertebrates to Gli1-3, through a Gli-consensus binding sequence, TGGGTGGTC (reviewed by Hooper and Scott, 2005; Jacob and Briscoe, 2003; Ruiz i Altaba et al., 2003). In the absence of Hh signaling in Hh-responsive tissues in mice, Gli3 and, to a far lesser extent, Gli2 (Pan et al., 2006) undergo proteosome-mediated carboxyl cleavage to repressor forms (Gli^{REP}) that silences one set of Hh targets. Hh signaling at low levels counters cleavage, leading to de-repression of targets. High levels of Hh signaling are essential for direct activation of a second group of targets. All three Gli members can function as activators (Gli^{ACT}), although only Gli1, itself a direct target of Hh signaling, is thought to act exclusively as an activator (Bai et al., 2004; Dai et al., 1999; Wang et al., 2000). The primary transcriptional activator appears to be Gli2; however, substitution of Gli1 into the Gli2 locus rescues the Gli2 null phenotype. Thus, Gli1 and Gli2 activator forms have similar

properties (Bai and Joyner, 2001). While not comprehensively explored, evidence from *Drosophila* suggests that Gli3 repressor and activator forms appear to bind a common set of target sites (Muller and Basler, 2000).

One crucial role of sonic hedgehog (Shh) lies in patterning of the vertebrate neural tube. Here, Shh secreted from the midline notochord acts as a morphogen to specify five classes of ventral neuronal progenitors in a concentration-dependent fashion; from dorsal to ventral these are progenitors for V0, V1, V2 interneuron pools, motoneurons (MN) and V3 interneurons. In addition, the highest levels of Shh signaling at the ventral midline induce a second ventral midline domain of Shh-expressing cells, the floor plate (FP), immediately ventromedial to V3 progenitors (reviewed by Briscoe and Ericson, 2001; Jessell, 2000; McMahon et al., 2003). Genetic studies have demonstrated that whereas V0, V1, V2 and MN identities can be specified in the absence of Hh signaling if Gli^{REP} is removed, the most ventral identities, V3 and FP, require Hh signaling and Gli^{ACT} forms (Bai et al., 2004). Recent work indicates that relatively small differences in the ratio of these forms dictate distinct transcriptional outputs (Stamataki et al., 2005), suggesting a fine-tuned response to Gli regulators in the cis-regulatory regions of target genes expressed at distinct Hh thresholds. Clearly, testing this or any alternative model requires a thorough understanding of the cis-regulatory mechanisms in direct target genes that dictate Gli-mediated regulation.

Here we have initiated a systematic effort to define sets of genes that are directly regulated by Gli activator. Gli1-directed chromatin immunoprecipitation (ChIP) products isolated during the course of Hh-mediated patterning of neural progenitors were used to interrogate custom genomic arrays, successfully identifying and mapping a large number of novel in vivo target sites. Bioinformatic analysis and expression studies in cell culture and transgenic mouse embryos validated these data and provide novel insights into the regulation of both tissue-specific and more general components of a Hh transcriptional response. Further, the analysis of in vivo targets facilitated the development of a novel algorithm for Gli-target site prediction. Based on our data, we present a model for ventral

[1]Department of Molecular and Cellular Biology and [2]Department of Statistics, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA. [3]Department of Statistics, Stanford University, Sequoia Hall, 390 Serra Mall, Stanford, CA, 94305, USA. [4]Agilent Technologies, 245 First Street, Suite 105, Cambridge, MA 02142, USA. [5]Institute for Systems Biology, Seattle, WA, 98103, USA. [6]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. [7]Harvard Stem Cell Institute, 16 Divinity Avenue, Cambridge, MA 02138, USA.

*Present address: Bioengineering Department, University of Illinois at Urbana-Champaign, 1304 West Springfield Avenue, Urbana, IL 61801, USA
†Author for correspondence (e-mail: mcmahon@mcb.harvard.edu)

neuronal specification in which Gli-activator and Gli-repressor forms differ substantially in their selection of binding sites. This uneven competition drives the transcriptional interpretation of the morphogen gradient.

MATERIALS AND METHODS
Generation of ES cells and transgenics
A full-length mouse Gli1 construct was cloned with an in-frame C-terminal 3XFLAG tag (Sigma). The tagged Gli induced luciferase expression 111±6-fold compared to 254±27-fold with the untagged form. This was cloned upstream of an IRES2, allowing bicistronic expression of the Venus YFP protein (Nagai et al., 2002), and the module was cloned into the pBigT shuttle vector, then into Rosa26PAS (Srinivas et al., 2001). The linearized construct was electroporated into YFP3-l (Rosa26YFP/β-gal) embryonic stem (ES) cells (Mao et al., 2005) and neomycin-resistant colonies that passed initial visual screens (loss of β-gal or YFP expression) were assayed by Southern blot. We used one ES cell line, Gli1[FLAG], for all further experiments. To obtain a constitutively active Rosa Gli1 construct (constitutive Gli1[FLAG]), the conditional ES cells were grown for several passages in 1 μm 4-OH Tamoxifen, and then cloned by serial dilution (three independent lines were isolated and we utilized one for all subsequent experiments). For leptomycin B treatment, ES cells were incubated for 5 hours with 5 nM leptomycin B.

The pHSP68lacZ2XINS vector was constructed by PCR amplifying the 2X Chick β-globin insulator motif from XB3+Insulator with primers containing Asc1 sites and inserting this into the Asc1 site of pHSP68lacZ (Sasaki and Hogan, 1996). With the exception of the Nkx2.1 (also known as Titf1 – Mouse Genome Informatics) peak, where only one copy of the enhancer was used, transgenic constructs were generated by cloning four copies of test DNA upstream of the minimal promoter in pHSP68lacZ2XINS. The sequence coordinates of the Ptch1 peak2 enhancer and the Nkx2.2 (Nkx2-2) Gli enhancers are described in the text. In addition, a 575 bp sequence (chr12:53,245,800-53,246,374) encompassing the Nkx2.1 peak and a 291 bp sequence (chr11:77915085-77915375) spanning the Rab34 peak were used to generate the respective constructs. To construct mutated Gli sites, we changed the site 5′-TGGGTGGTC-3′ to 5′-TCCCACGTC-3′ (Nkx[M]-HSP68lacZ), a mutant form that removes an invariant G essential for Gli binding and several other residues that contact the zinc finger motif of Gli1 (Pavletich and Pabo, 1993).

Embryoid bodies (EBs) were generated using previously described procedures (Wichterle et al., 2002). After 2 days, the media was changed to DFNB containing 500 nM retinoic acid and, when applicable, 1 μM Hh agonist (HH-Ag) (Frank-Kamenetsky et al., 2002). EBs were grown for an additional 3 days to induce neural progenitor stages, at which point cultures were harvested for ChIP.

Immunostaining with Pax6, Nkx6.1 (Nkx6-1), Nkx2.2 and FoxA2 was performed using previously published methods (Wijgerde et al., 2002). For western blots, samples were run out on 4-12% BisTris gradient gels and protein levels were normalized using levels of β-gal. The M2 FLAG antibody (Sigma) or anti-β-galactosidase (Promega Z3781) were each used at a 1:2000 dilution. For immunostains, YFP was detected using an anti-GFP at 1:2000 (Abcam ab290).

Chromatin immunoprecipitation and hybridization
We modified a published protocol (Odom et al., 2004) to reduce the number of input cells to approximately 2×10^6 cells per precipitation using a FLAG M2 mouse monoclonal antibody (Sigma). After isolating chromatin, we incubated it with previously prepared M2-coated magnetic beads. To prepare beads, magnetic sheep anti-mouse IgG beads (Dynal) were incubated overnight with 0.8 μg M2 Mouse monoclonal anti-Flag antibody (Sigma) and 10 μl of beads per precipitation. The following day, the beads were rinsed and added to chromatin and incubated overnight at 4°C. Samples were then rinsed five times with RIPA buffer, and the antibody was stripped from the beads by incubating in 1% SDS at 65°C for 15 minutes and cross-linking was reversed by incubating overnight at 65°C. The next day, samples were sequentially treated with RNAse A and Proteinase K, phenol:chloroform extracted, ethanol precipitated with 20 μg glycogen and resuspended in 60 μl 10 mM Tris pH

8.0. ChIP-enriched DNA samples were amplified before hybridization to the tiling array. For this, DNA samples were blunted using T4 DNA polymerase (50 μl of each ChIPed sample with 0.6 units of T4 DNA polymerase in a 110 μl volume at 12°C for 15 minutes), phenol:chloroform extracted with 10 μg glycogen, ethanol precipitated and resuspended in 25 μl water. This material was then blunt-ligated to unidirectional linkers (Ren et al., 2000) using 15 μM annealed linkers and a Quick Ligation Kit (New England Biolabs) in a 21 μl volume at 12°C overnight. The reaction was then ethanol precipitated, resuspended in 25 μl water and PCR amplified using 1 μM primer oJW102 with an annealing temperature of 60°C and a 1 minute extension at 72°C for 26 or 29 cycles. To ensure that samples were uniformly amplified, we re-assayed the samples by qPCR.

For each ChIP, 2 μg amplified ChIP sample or Input control was labeled with Cy3 or Cy5-dUTP (Perkin Elmer) using the Bioprime Array CGH kit (Invitrogen) and competitively hybridized using 5 μg labeled material per channel with the Oligo CGH Hybridization Kit (Agilent Technologies) on a custom tiled 44K array from Agilent Technologies (see Table S1 in the supplementary material). Complete details for array construction and all raw array data may be obtained from GEO (GEO #GSE5683). Three independent biological samples were employed in the analysis. Within each sample, dye-swap experiments were performed in duplicate for a total of four technical replicates per biological sample.

The hybridizations were done in Agilent SureHyb hybridization chambers, and in an Agilent rotating hyb oven at a speed of 10 rpm for 72 hours at 65°C. Each hybridization was performed in duplicate and by exchanging labeled dyes (dye-swapping) for a total of four technical replicates per biological sample. Samples were washed sequentially for 5 minutes at room temperature using Oligo aCHG Wash Buffer 1 (Agilent Technologies), 1 minute at 37°C in Oligo aCGH wash buffer 2 (Agilent Technologies), 1 minute in Acetonitrile at room temperature and 30 seconds in Stabilization and Drying Solution (Agilent Technologies). Finished arrays were scanned using an Agilent scanner at a 10 μm resolution with both channels at a PMT setting of 100. Array images were extracted using Agilent's Feature Extraction Software (version 8.5.1.1) using the standard CGH protocol included with the software. The settings included spatial detrending of the extracted array data and a linear median normalization. Log fold changes were obtained for each probe, and the four technical replicates within each biological replicate were averaged. The correlation coefficient between any two technical replicates using the same dye labeling was $r=0.948$ (s.d.=0.020), and between two technical replicates that use different dye labeling was $r=0.773$ (s.d.=0.095). In contrast, the correlation between two biological replicates was $r=0.068$ (s.d.=0.075). T-scores were obtained using TileMap in the moving average mode (Ji and Wong, 2005). All coordinates in this manuscript refer to the Build 34 assembly of the mouse genome.

Mapping of Gli sites, determination of conserved regions and motif identification
Potential binding regions (see Table S2 in the supplementary material) were extended 200 bp from both ends and repeats were masked (A. F. A. Smit, R. Hubley and P. Green, RepeatMasker at http://repeatmasker.org). We then mapped the Gli consensus-binding pattern TGGGTGGTC (Kinzler and Vogelstein, 1990), allowing only one bp mismatch. Candidate binding regions were binned into groups of size ten according to their initial rank, and the number of Gli sites (n_1) and the total number of surveyed non-repeat base pairs (n_2) were counted for each tier and reported in Table S3 in the supplementary material. To determine the background level of Gli occurrence, the Gli consensus sites were also mapped to all tiled regions present on the array, yielding 1264 sites in a total of 5,608,267 bp non-repeat sequences. Gli-binding-site enrichment was then computed for each tier by comparing its site density to the site density of control regions, i.e. $r_1=(n_1/n_2)_{target}/(n_1/n_2)_{control}$.

In a separate approach, we examined the degree of cross-species conservation among binding sites. The multiple species alignment among mouse, rat, human, dog and zebrafish were used to compute a conservation score for each position of mouse genome. Sites whose mean conservation score is among the top 10% of the genome were called conserved sites. The

Table 1. Summary statistics of ChIP peaks

Rank	Gene	Peak	Chr.	Start	End	qPCR	Orientation	Gli concensus	Gli matrix
1	*Ptch1*	2	13	60949397	60950371	34	TSS –1.1k	–137, 219	–154, –137
2	*Nkx2-2*	1	2	146644626	146645570	29.2	TSS –1.9k	–51	–98, –51
3	*Ptch1*	3	13	60951377	60952953	45.3	TSS –3.7k	78, 437	–760, 78
4	*Nkx2-9*	2	12	53347310	53348354	16.2	TSS –8.7k	–25	–222, –25, 45, 289
5	*Ptch1*	5	13	60956284	60956777	65	TSS –7.9k	–78	–78
6	*Nkx2-9*	1	12	53340519	53341015	37.7	TSS –1.9k	–119	–119
7	*Rab34*	1	11	77915220	77915560	28.1	Intron 1,2	–71, 264	–106, –71
8	*Ptch2*	1	4	116057320	116057946	5.6	Intron 2	–106	–106
9	*Ptch1*	1	13	60945578	60945972	30	Intron 2	101	
10	*Nkx 2-1*	1	12	53245748	53246351	9.1	TES +11.5k	152	–355, 152
11	*Gli1*	1	10	127076571	127078504	7.2/4.7/12.5	Promoter, Intron 1,2	–115, 543, 1422, 1580	–115, 1580
12	*Hhip*	1	8	79270092	79270492	5.6	Intron 1	–51	–184, –51
13	*FoxA2*	1	2	147493701	147494245	6.4	TES +5.8k	–148, 331	–148, 331
14	*Ptch2*	2	4	116054385	116054954	48.3	TSS –0.3k	–136, 61	–136, 65
15	*Cart1*	1	10	102925949	102926177	–	Intron 3	NA	NA
16	*Ptch1*	4	13	60953950	60955072	9	TSS –6.3k	–36	380
17	*Prdx2*	1	8	84215218	84215559	1.4	TSS –22.5k	NA	–78
18	*Cart1*	2	10	102938108	102938308	–	Intron 2	–75	–75
19	*Ptch1*	6	13	61010241	61010714	2.8	TSS –61.9k	NA	NA
20	*Hhip*	2	8	79268419	79268867	8.9	Intron 1	NA	NA
21	*Flrt3*	1	2	140191468	140191692	–	TSS –6.3k	NA	NA
22	*Pax9*	1	12	53443055	53443311	–	TES +5.3k6	–35	NA
23	*Ncor2*	1	5	124340743	124341009	–	TSS –0.4k	60, 335	54, 60,199
24	*Zic3*	1	X	52789693	52789967	–	TES –0.8k (3′ UTR)	NA	NA
25	*Hand2*	1	8	56373391	56373692	2.7	TES +10.3k	–154	–181, –154

First-tier peaks (ranked 1-13) include all known Gli enhancer sites (gene name is underlined in 'Gene' column). Negative values indicate regions upstream (5′) of the specified reference site in the orientation column, and positive values are downstream (3′) of the site. Gli consensus and matrix sites were mapped to the defined peak regions and are displayed as values relative to the peak center (see Table S2 in the supplementary material). Our Gli-site mapping tolerates only a 1 bp mismatch from the TGGGTGGTC consensus. Sites that are identified by both consensus and matrix mapping are underlined. Chr., chromosome; NA, not applicable (no Gli sites present); TSS, transcriptional start site; TES, transcriptional end site; UTR, untranslated region.

number of conserved Gli sites (n_3) and the total number of surveyed base pairs that are conserved (n_4) were counted for each tier and control regions. We computed the enrichment of conserved Gli sites in conserved regions as:

$$(1)\ r_2 = (n_3/n_4)_{target}/(n_3/n_4)_{control};$$

and

$$(2)\ r_3 = (n_3/n_2)_{target}/(n_3/n_2)_{control}.$$

r_2 measures the extent to which Gli sites are enriched in conserved regions, and r_3 measures the combined gain we can obtain using both binding-sequence specificity and cross-species conservation information.

To perform de novo motif discovery, Gibbs motif sampler was applied to the top 13 tier 1 peaks after extending each defined peak 100 bp from both ends. The Gli matrix identified by motif analysis were visualized using WebLogo (Crooks et al., 2004). To map Gli consensus sites, the core-binding pattern TGGGTGGTC was used to scan genomic sequences; up to one mismatch was allowed. We computed cross-species conservation scores using mouse, rat, human, dog and zebrafish alignments. Motif sites that reside within the top 10% most conserved regions in the mouse genome are defined as phylogenetically conserved sites.

In silico predictions of Gli target genes and enhancers
To generate an experimental dataset for predictions, retinoic acid-treated EBs were grown in the presence or absence of HH-Ag. We did not observe induction of ventral Hh markers in RA-treated constitutive Gli1[FLAG] EBs and used these for the control, baseline set. To pick genes that were upregulated by Hh signaling, a modified t-statistic (Ji and Wong, 2005) was computed for each gene to test if its expression level is higher in Hh-induced samples compared with non-induced samples and the top 365 probe sets corresponding to 249 non-redundant genes were used in further analyses (see Table S5 in the supplementary material). We then identified their orthologs from human, dog, cow, chicken and zebrafish. Each gene was extended 50 kb upstream from transcription start and 50 kb downstream from transcription end, and predictions were made based on these sequences.

In order to make EEL predictions, EEL was run using the same parameters as Hallikas et al. (Hallikas et al., 2006). In addition to the previous 107 motif-position-specific weight matrices, we also included the Gli matrix recovered in this current study (see Fig. S2 in the supplementary material). Pairwise alignments between mouse-human, mouse-dog, mouse-cow, mouse-chicken and mouse-zebrafish were generated. As EEL did not provide a way to combine these pairwise alignments into multiple alignments and to rank predicted enhancers accordingly, we decided to use mouse-human alignments as the basis for making our predictions. Each predicted enhancer was then annotated to reflect whether it was also found in other pairwise alignments. Predictions shorter than 2000 bp, with an EEL score $\geqslant100$ and a combined Gli score $\geqslant25$ were picked, retained and ranked by their combined Gli scores.

To perform MCA, we first collected conserved, non-coding genomic segments from the mouse genome. Cross-species conservation scores were computed based on percent identities measured in a 50 bp sliding window using multiz mouse, rat, human, dog and zebrafish alignments (downloaded from http://genome.ucsc.edu). The scores ranged from 0 to 255, with 255 corresponding to the most conserved status. We collected all continuous segments with a score $\geqslant40$ (top 10%), discarding segments shorter than 50 bp. These first-stage segments were joined to form second-stage segments if their end-to-end distance (gap) was $\leqslant50$ bp. The second-stage segments were filtered further to remove segments <200 bp or if they did not contain at least one first-stage segment >100 bp. The remaining segments were termed third-stage segments, comprising the final collection of conserved genomic regions.

To measure Gli enrichment, we mapped the Gli matrix recovered from the ChIP to all third-stage conserved genomic regions. The Gli matrix was compared to a third-order background Markov model, and a likelihood ratio $\geqslant500$ was used as a cutoff to define Gli sites. The occurrence of Gli sites in conserved genomic segments was then modeled as Poisson processes. If a segment is a Gli-binding region (the alternative hypothesis H_1), the occurrence rate of Gli sites was assumed to be λ_1 (site/bp) and each Gli site was assumed to be generated from the Gli-binding matrix. In contrast, if a segment is not a Gli-binding region (the null hypothesis H_0), the occurrence rate of Gli sites was assumed to be λ_0 and all such sites were assumed to be

Table 2. Global versus tissue-restricted enhancers

Region	Coordinates	Size (bp)	Conserved Gli sites	*Gli1* induction	St. dev.
Ptch1 peak 1	chr13:60,945,834-60,946,202	369	chr13:60946069-60946077(–)	13.5	2.3
Ptch1 peak 2	chr13:60949285-60949464	180	chr13:60949424-60949435(–)	11.2	3.7
Ptch1 peak 3	chr13:60952107-60952356	250	chr13:60952273-60952284(+)	3.3	0.8
Ptch1 peak 4	chr13:60,954,478-60,955,071	594	chr13:60954710-60954718(–) /// chr13:60954791-60954799(–)	3.8	0.4
Ptch1 peak 5	chr13:60956197-60956431	235	chr13:60956316-60956327(+)	2.6	0.5
Ptch1 peak 6	chr13:61,010,209-61,010,707	499	–	3.1	1.3
Ptch2 peak 1	chr4:116,057,081-116,057,672	592	chr4:116057540-116057551(+)	7	2.1
Ptch2 peak 2	chr4:116,054,618-116,055,117	500	chr4:116054800-116054808(+) /// chr4:116054812-116054823(+) /// chr4:116054906-116054914(+)	129	24
Hhip1 peak 1	chr8:79,269,855-79,270,597	743	chr8:79270163-79270171(+) /// chr8:79270296-79270304(–)	7.7	2.2
Hhip1 peak 2	chr8:79,268,170-79,268,855	686	–	1.8	0.7
Nkx2.2	chr2:146645052-146645290	239	chr2:146645185-146645196(+)	3.2	1
Negative control	chr12:53255352-53255521	170	–	2.6	0.6
Rab34	chr11:77915199-77915431	233	–	6.2	1.1
Empty vector	–	–	–	2.1	0.5

Conserved Gli motifs in ChIPed regions near predicted global responders to Hh signaling were used to obtain putative Gli enhancer elements that were cloned into a minimal promoter luciferase vector and transfected into murine NIH3T3 fibroblasts in the presence or absence of Gli1. We detect no specific induction for the neural-specific Nkx2.2 peak, suggesting that this assay only identifies global responders of Hh signaling. chr, chromosome; St. dev., standard deviation.

false sites and generated from the background Markov model. The rate λ_0 was estimated based on the mapping results in general conserved regions (λ_0=0.0002/bp), and λ_1 was estimated using conserved regions covered by the top 25 peaks (λ_1=0.0015/bp). For each segment, the log likelihood ratio between H_1 and H_0 was computed as:

$$S = n \times \log_{10}(\lambda_1/\lambda_0) - (\lambda_1-\lambda_0) \times l \times \log_{10}(e) + \log_{10}(L_1) - \log_{10}(L_0).$$

Here, l is the length of a conserved segment, n is the number of Gli sites in the segment obtained by matrix mapping, L_1 is the probability to generate the sites according to Gli matrix, and L_0 is the probability to generate the sites by background Markov model. Finally, S was used to measure the Gli enrichment and rank conserved segments.

We also mapped Gli consensus TGGGTGGTC (allowing $\leqslant$1 mismatch) to the third-stage conserved genomic regions. All regions that contain at least one Gli consensus site were selected as a potential Gli cis-regulatory module (CRM). The Gli CRMs are then ranked by Gli enrichment score S. For each gene, all Gli CRMs located within introns, UTRs, 50 kb upstream regions (from transcription start) and 50 kb downstream regions (from transcription end) were collected and the enrichment scores were added together to derive the combined Gli-binding strength for that gene.

Cell culture
Gli luciferase assays in murine NIH3T3 cells used methods described previously (Nybakken et al., 2005). We used 25 ng Gli1 cloned into pCIG (Megason and McMahon, 2002) or empty pCIG vector for controls. Candidate enhancers (Table 2) were PCR-amplified from genomic DNA and cloned into the pGL3-Promoter vector (Promega).

Transcriptional profiling
To generate the list of Gli target candidate genes, a variety of mouse stages and Hh-pathway mutants (*Ptch1*, *Smo* and *Shh*) conditions were analyzed by transcriptional profiling (T.T. and A.P.M., unpublished). Briefly, for each combination, three different samples were profiled using a combination of Affymetrix mouse U74Av2 and MOE430 A&B arrays. The screen included 6-8-somite embryo (E8.5), 10-13-somite embryo (E8.75), head (E10.5) and limb (E10.5). Data were analyzed by dChip (Li and Wong, 2001) and PowerExpress (H.J. and W.H.W., unpublished) to identify and rank putative Hh targets. The EB transcriptional profiling screen was conducted by screening quadruplicate samples of neuralized EBs treated with Hh agonist (Frank-Kamenetsky et al., 2002) and 500 nM retinoic acid (under identical conditions to the ChIPs) versus EBs treated with 500 nM retinoic acid alone. Samples were profiled using the Affymetrix MOE4302.0 arrays and results were analyzed using dChip (GEO #GSE4936).

RESULTS
Gli1 ES cell line and differentiation into neuralized embryoid bodies
Our analysis focused on Gli1, the only Gli member that appears to function exclusively as an activator in Hh signaling. In order to circumvent the lack of useful Gli antibodies, we generated a biologically active C-terminal FLAG-tagged Gli1 construct. Gene targeting introduced this into the constitutive *ROSA26* locus such that Cre-mediated excision of a selection cassette was required to activate Gli1[FLAG] production (see Fig. S1A in the supplementary material). The Cre-dependent inducibility of the Gli1[FLAG] construct in ES cells was confirmed by immunostaining and western blot analysis (see Fig. S1B,D in the supplementary material) and a clonal line was established in which Gli1[FLAG] was constitutively produced (see Fig. S1C in the supplementary material). As previously reported, Gli1[FLAG] was predominately restricted to the cytoplasm but accumulated rapidly in the nucleus when nuclear export was inhibited by the addition of leptomycin B (Kogerman et al., 1999) (see Fig. S1C in the supplementary material).

As a strategy for isolating Gli1 targets in the ventral neural tube, we focused on the demonstrated ability of Shh to recapitulate its ventral neuralizing activity in the specification of ES cells differentiated to form neuralized EBs (Wichterle et al., 2002). In this assay, addition of Shh, or small molecule agonists that activate the Shh pathway, led to a quantitative ventralization of neural progenitors closely mirroring the neural patterning activity of Shh in the ventral neural tube. In the absence of Hh agonist (Hh-Ag), the Gli1[FLAG] line was not sufficient to specify Nkx6.1[+]- (V2, MN, V3, FP progenitors), Nkx2.2[+]- (V3 progenitors) or FoxA2[+]- (FP) producing cell types (Fig. 1B-D,J-L). The only differential response detected between the induced Gli1[FLAG] line and control, parental EBs was a large decrease in the number of Pax6[high] cells, although the total number of Pax6-producing cells was not altered (70% σ=10% of cells are Pax6[high] in control EBs compared with 2% σ=4% for Gli1[FLAG] EBs; Fig. 1A,I,Q); Pax6 repression has been shown to be Shh-dependent (Ericson et al., 1997). In contrast to this weak response, Gli1[FLAG] resulted in a strong synergy with Hh-Ag (Frank-Kamenetsky et al., 2002). At a dose of 1 μM, EBs were only weakly ventralized (Nkx6.1[+], Nkx2.2[–], FoxA2[–]). In contrast, when Gli1[FLAG] was present, an identical dose of Hh-Ag resulted in the
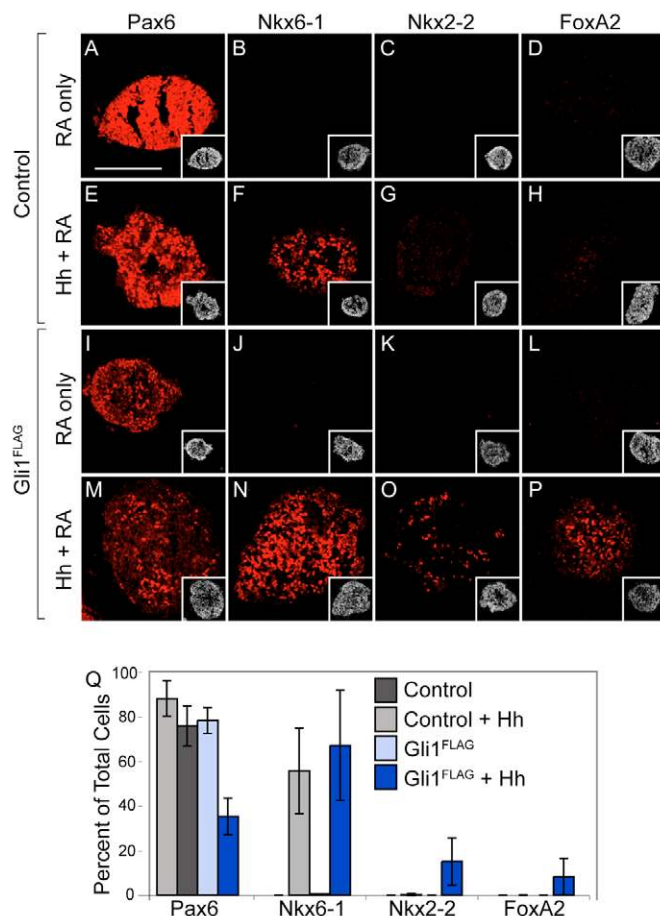
Fig. 1. EBs containing Gli1^FLAG exhibit an amplified response to Hh stimulation. (**A-P**) Immunostaining of EBs to detect neural progenitor types in mouse. While both control (A-H) and Gli1^FLAG samples (I-P) respond to Hh signaling by activating Nkx6.1 expression (F,N), only the latter samples contain Nkx2.2^+ and FoxA2^+ cells on Hh-Ag treatment (O,P). (**Q**) Distinct neural progenitors demarcated by expression of progenitor-type-specific markers were counted and represented as a percentage of all DAPI-positive cells in three independent EBs. The error bars represent the standard deviation. Scale bar: 100 μm.

induction of ventralmost neural (Nkx2.2^+, V3 progenitor) and FP (FoxA2^+) identities (compare Fig. 1F-H with N-P), as well as a substantial reduction in Pax6^+ cells consistent with Nkx2.2 repression of Pax6 (Fig. 1E,M) (Ericson et al., 1997). Nkx2.2^+ and FoxA2^+ cells comprised approximately 10% each of the EB-derived population. The requirement of Shh for Gli1 activation – even when under control of an ectopic promoter – was previously noted by Bai and Joyner (Bai and Joyner, 2001).

Chromatin immunoprecipitation and statistical validation of putative Gli-binding sites

In an independent study, the developing neural tube and several additional Hh target populations in the early mouse embryo were transcriptionally profiled in an attempt to identify possible targets (positive and negative) of Hh signaling (Tenzen et al., 2006) (T.T. and A.P.M., unpublished). We ranked these genes using PowerExpress (http://biogibbs.stanford.edu/~jihk/CisGenome/microarray.htm) (H.J. and W.H.W., unpublished) and chose the top

122 genes to place on the custom tiling array. For each gene, this method uses the TileMap probe level hierarchical model to compute a posterior probability that the gene has a positive or negative Hh-responsive expression pattern. Among the top targets on this list were several general Hh pathway components, including *Ptch1*, *Gli1*, *Hhip1* and *Gli3*, and previously reported direct neural (*FoxA2*) (Sasaki et al., 1997) and mesodermal (*Myf5*) (Teboul et al., 2003) targets of Hh signaling. A total of 122 genes were selected, together with approximately ten additional genes drawn from the literature to generate arrays for ChIP-mediated identification of Gli1 targets (see Table S1 in the supplementary material). The inclusion of non-neural targets provides an important control for context-dependent regulation in these studies. Custom arrays were generated after repeat masking in which genomic sequence encompassing 25 to 75 kb 5′ and 3′ of the transcriptional start site (TSS) was represented by a 60-mer DNA probe at a spacing of one 60-mer every 125 bp for a total of approximately 5.6 mb of surveyed genomic sequence. Because of the large variability in the size of a given transcriptional unit, probes covered an extensive region 5′ to the TSS, but a variable extent of intronic and/or other 3′ regions.

To verify Gli1^FLAG-specific enrichment of expected cis-regulatory regions in Hh-Ag-induced EBs, we adapted a standard ChIP protocol to perform ChIP on 2×10^6 cells (Odom et al., 2004). Chromatin extracts were incubated with anti-FLAG antibody and following IP, the enrichment of known Gli targets in IP fractions was examined by quantitative RT-PCR (qPCR). qPCR confirmed specific enrichment of Gli-binding sites in several known target regions, including the *FoxA2* FP enhancer (Sasaki et al., 1997), a conserved cis-regulatory region upstream of *Ptch1* previously associated with Shh-dependent regulation in human *Ptch1* (Agren et al., 2004), and *Gli1* itself (Dai et al., 1999). *Gli1* expression is absolutely dependent on prior Hh signaling (Bai et al., 2004).

Having verified the ChIP procedure, we screened the custom tiling array. The regions significantly enriched in the IP fraction were ranked by a T-score, identifying 47 peaks having a false discovery rate of ≤25% (see Table S2 in the supplementary material). As the top 13 ranked sites contained all previously known Gli targets known to be expressed in the neural tube (Table 2, underlined), we used these regions to determine if we could uncover any enriched motif that might represent a transcription-factor-binding site without any prior knowledge of that site. Employing a Gibbs motif sampler (Lawrence et al., 1993; Liu, 1994), we obtained two distinct motifs. The first motif (Fig. 2I) was very similar to the Gli-binding motif (TGGGTGGTC; Fig. 2J; see Fig. S2 in the supplementary material) determined by oligonucleotide binding to recombinant proteins (Hallikas et al., 2006; Kinzler and Vogelstein, 1990; Vortkamp et al., 1995), a confirmation that our strategy specifically isolates sequences directly bound by Gli1^FLAG. Further, the analysis predicted nucleotide preferences extending beyond the core Gli consensus region (Fig. 2I). A second, lower scoring motif with a GC-rich pattern was the only other significantly enriched sequence; this is not enriched in conserved regions and is also present in multiple, unrelated ChIP datasets, suggesting that it is non-specific (data not shown).

In an initial effort to determine the biological significance of these regions, we used the enrichment and phylogenetic conservation of Gli motif sites as a measurement of the likelihood that a given peak represents specific Gli1 binding. The ranked sites were grouped into bins of ten, the Gli consensus motif TGGGTGGTC (≤1 mismatch) was enriched in peak regions with high T-scores, when compared
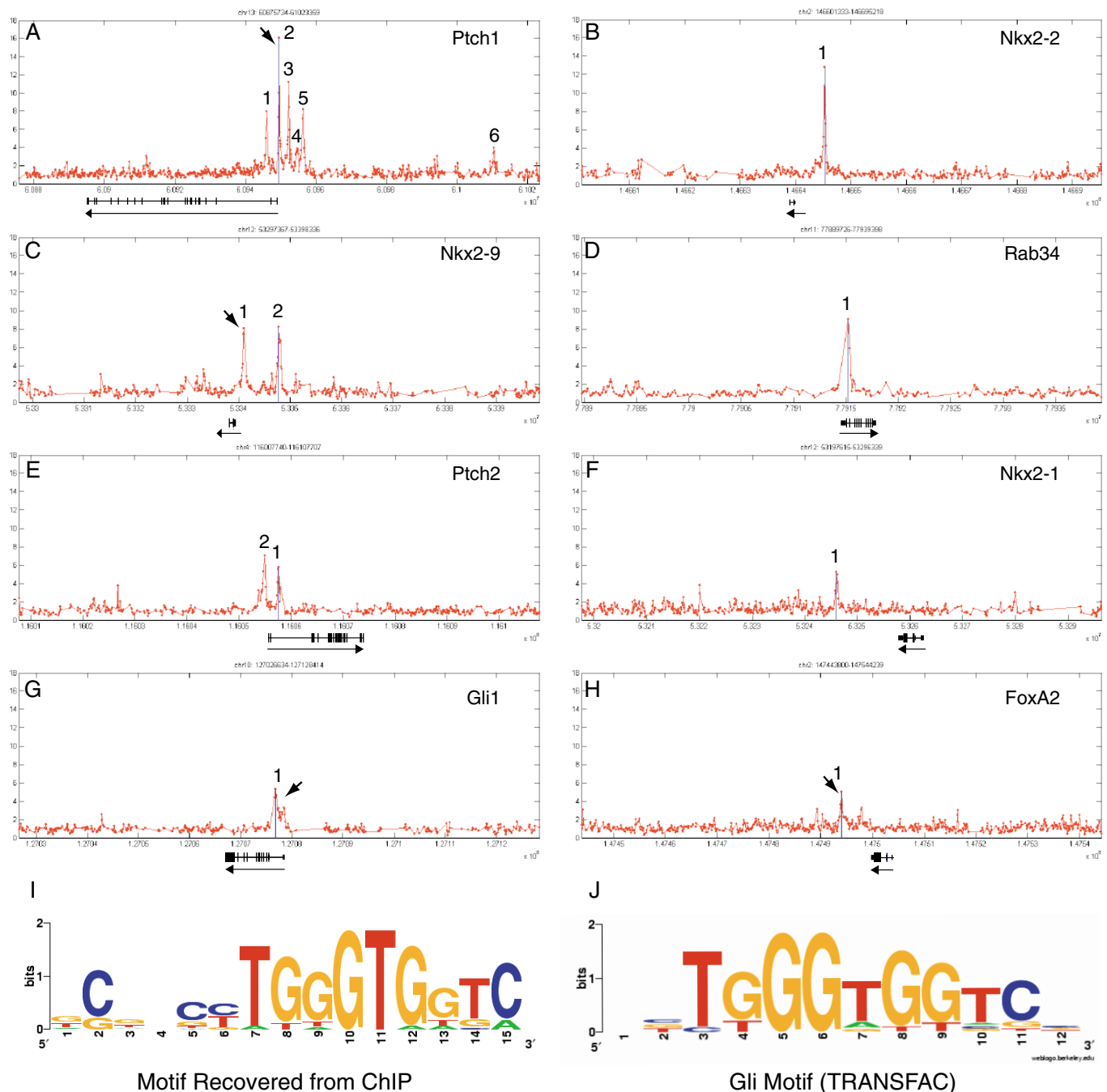
Fig. 2. Chromatin immunoprecipitation of neuralized murine EBs. (**A-H**) Selected positive targets showing the mean fold enrichment of Gli1[FLAG] ChIPs representing three biological replicates, each with four technical replicates. The plots show the mean fold enrichment of ChIPed sequence versus input. Approximately half of the peaks lie outside the proximal promoter region. Multiple peaks are numbered to correspond to the peaks listed in Table 1 and in Table S2 in the supplementary material; arrows indicate previously described Gli regulatory regions. Below each graph, the position of exons (rectangles) and the direction transcription (arrows) are shown relative to the ChIPed region. (**I,J**) De novo motif analysis recovers a consensus site (I) similar to a previously described Gli1 consensus sequence (J) (Transfac # M01037) (I).

with control regions; however, we observed a distinct drop in enrichment from about 5-fold in the first two bins to 3.5-fold in the third bin (see Table S3 in the supplementary material). Of the five motif sites detected in the third bin, four of them occurred in regions ranked 21-25, and only one in regions ranked 26-30. Consequently, we focused on the top group of 25 predictions in subsequent analyses. Fig. 2 shows examples of the data for a subset of genes studied here. These 25 regions contained 28 Gli consensus motif sites in a total of 23,656 bp of non-repeat sequences. The width of the peak itself is predicted to depend both on the number of Gli sites within a particular region and the length of chromatin fragments generated on DNA shearing. The first, second (median) and third quantiles of region length were 295, 474 and 707 bp, respectively. Many, but not all, peaks are centered about a strong Gli consensus

sequence. In total, we observed a 5.25-fold enrichment of predicted Gli target sites within peaks when compared with all regions tiled on the array.

If we were to assume that the top 25 ChIP peaks represented all the Gli-binding events on the array, only 28/1264 Gli consensus sites (2.22%) were bound by Gli (some ChIP peaks had multiple consensus sites, while others did not contain any consensus sites). Among the 1264 consensus sites tiled in the array, 299 were phylogenetically conserved, and 20 of these (6.69%) were associated with Gli binding. These regions were associated with 16 genes, indicating that a significant fraction of the targets contained multiple Gli-binding modules. Interestingly, only seven of the regions identified (28%) were within the 5 kb proximal promoter region. Remaining sites were located in the first or second intron at varying distances from the TSS (24%), at a range greater than 5 kb 5′ of the TSS (24%), after the end of transcription (TES) (16%) or elsewhere within the transcript (8%) (Table 1). These results highlight the problem of local promoter analyses for large-scale mining of transcriptional regulatory mechanisms in the mammalian genome.

To verify that ChIP-chip screening reflects enrichment of a given region of DNA before amplification, we assayed most of the top 25 regions by qPCR using unamplified ChIP material and input chromatin (pre-ChIP baseline) that was a technical replicate of one of the amplified samples. When compared with baseline negative control values (see Table S4 in the supplementary material), all of the peaks present in the top half showed significantly higher qPCR signal in the ChIPed samples. In contrast, this was true for only half of the assayed peaks in the second half of the ranking (3/6; Table 1). Some of these peaks (i.e. *Hhip1* peak2) lacked a Gli matrix site but still exhibited a very significant enrichment by qPCR (Table 1), suggesting that other mechanisms, not simply direct DNA binding, might bring Gli factors to their target sites (see Discussion). As an additional specificity control, analysis of ChIP products in the parental ES cell line before Gli1^{FLAG} activation failed to show any enrichment, indicating that the ChIP results were indeed Gli1^{FLAG}-dependent (data not shown).

Identification of neural Gli enhancers
We next examined this list for the presence of known Gli-binding sites that mapped within reported cis-regulatory regions in *Ptch1* (Agren et al., 2004; Hallikas et al., 2006), *Gli1* (Dai et al., 1999), *Nkx2-9* (Santagati et al., 2003) and *FoxA2* (Sasaki et al., 1997); each of these predicted target sites was identified (arrows in Fig. 2A,G,C,H, respectively). To the best of our knowledge, these represent all published, mapped Gli-binding sites in targets expected to be expressed within the EB-generated ventral neural target population. Based on the presence of these previously validated sites, we further divided the ranked list into a first tier of 13 regions that included all the previously characterized Gli enhancer sites and a second tier of 12 regions that represented additional peaks of high statistical quality (Table 1, Table S2 in the supplementary material, and data below). As expected from our previous transcriptional profiling data, the transcripts adjacent to these peaks were upregulated in *Ptch1$^{lacZ/lacZ}$* embryos (Goodrich et al., 1997) and downregulated in *Smo$^{-/-}$* embryos (Zhang et al., 2001), where Hh signaling was enhanced or absent, respectively (see Fig. S3 in the supplementary material). In addition to the one previously reported Gli enhancer in *Nkx2.9* (Santagati et al., 2003) (Fig. 2C, peak 1), we detected an additional site (Fig. 2C, peak 2) approximately 8.7 kb 5′ to the TSS. In human *Ptch1*, a single 2981

bp enhancer/promoter fragment has been reported from studies in NIH3T3 and HEK293T cells; two conserved, predicted Gli-binding sites are predicted in this region, one of which has been shown to be functional (Agren et al., 2004). The homologous region was detected in our analysis (peak 2, arrow in Fig. 2A). We also identified five additional peaks, including three that were highly enriched (Fig. 2A). One of these, peak 5, lies immediately upstream of an uncharacterized alternatively spliced full-length form of *Ptch1*. While human *Ptch1* contains three alternatively spliced full-length transcripts (Agren et al., 2004), there are only two full-length forms represented among mouse expressed sequence tags (ESTs). A peak immediately upstream of *Gli1* is unusually broad, possibly reflecting an extended region of *Gli1* that also binds Gli3 (Fig. 2G) (Dai et al., 1999). We also detected strong peaks in *Ptch2* (Fig. 2E, two peaks), *Hhip1* (data not shown, two peaks), *Nkx2.2* (Fig. 2B, one peak), *Nkx2.1* (Fig. 2F, one peak) and *Rab34* (Fig. 2D, one peak); with the exception of *Nkx2.2* (Lei et al., 2006), none of these putative Gli-regulatory regions have been previously reported (Fig. 2). *Ptch2* and *Hhip1* are members of the Hh pathway that have been reported to show Hh-dependent regulation (Chuang and McMahon, 1999; Motoyama et al., 1998); *Nkx2.2* encodes a transcriptional determinant of the Shh-dependent V3 interneurons (Briscoe et al., 1999); *Nkx2.1* encodes a Shh-dependent transcriptional regulator within ventral neural populations of the forebrain (Pabst et al., 2000); and *Rab34* encodes a small GTP-binding protein not previously associated with Hh signaling that is predicted from our transcriptional array data to be a rather general target of Hh regulation in several tissues.

Differences between generic and tissue-restricted Gli enhancers
The genes identified as targets in our analysis exhibited either a global (multiple Hh-responsive tissues) or tissue-restricted pattern of expression (see Fig. S3 in the supplementary material). To test Gli-mediated regulation in a putative cis-regulatory enhancer region, we identified 12 candidate enhancers (Table 2) and asked if these could respond to Gli1 stimulation. These assays were performed in the presence and absence of Gli1 using murine NIH3T3 cells, a fibroblast line that has been used to define several cis-regulatory regions, including a *Ptch1* enhancer region (*Ptch1*, peak 2) identified in this screen (Agren et al., 2004). We focused our analysis of candidate sites among global Hh responders (*Ptch1*, *Ptch2*, *Hhip1* and *Rab34*) and compared it with the putative regulatory region in *Nkx2.2*, the expression of which has only been shown to require Hh signaling in the context of neural progenitor specification (see Fig. S3B-B″ in the supplementary material). As expected, we did not detect any specific Gli-mediated induction for the neural-specific *Nkx2.2* candidate enhancer in the reporter assay (Table 2). Nor did we observe reporter activation for either of the peaks in *Ptch1* (Fig. 2A, peak 6) or *Hhip1* (peak 2, data not shown) that failed to show a consensus Gli motif. However, we observed a robust induction with the Gli site in *Ptch1* peak 2 (Fig. 2C), as well as in previously unknown sites in *Ptch1* Intron2 (Fig. 2C, peak 1), and *Ptch2* (Fig. 2E, peak 2). A more modest response was detected in *Ptch2* peak 1 (Fig. 2E), *Hhip1* peak 1 (data not shown) and *Rab34* (Fig. 2D). Taken together, these data (Table 2) indicate that many Gli-responsive elements are confirmed in this assay; those that are not may reflect the absence of context-dependent factors (see Discussion).

To further explore the properties of a subset of these putative enhancers in more detail, we selected four of the regions for more extensive transgenic analysis. We chose a 180 bp Ptch1 peak 2
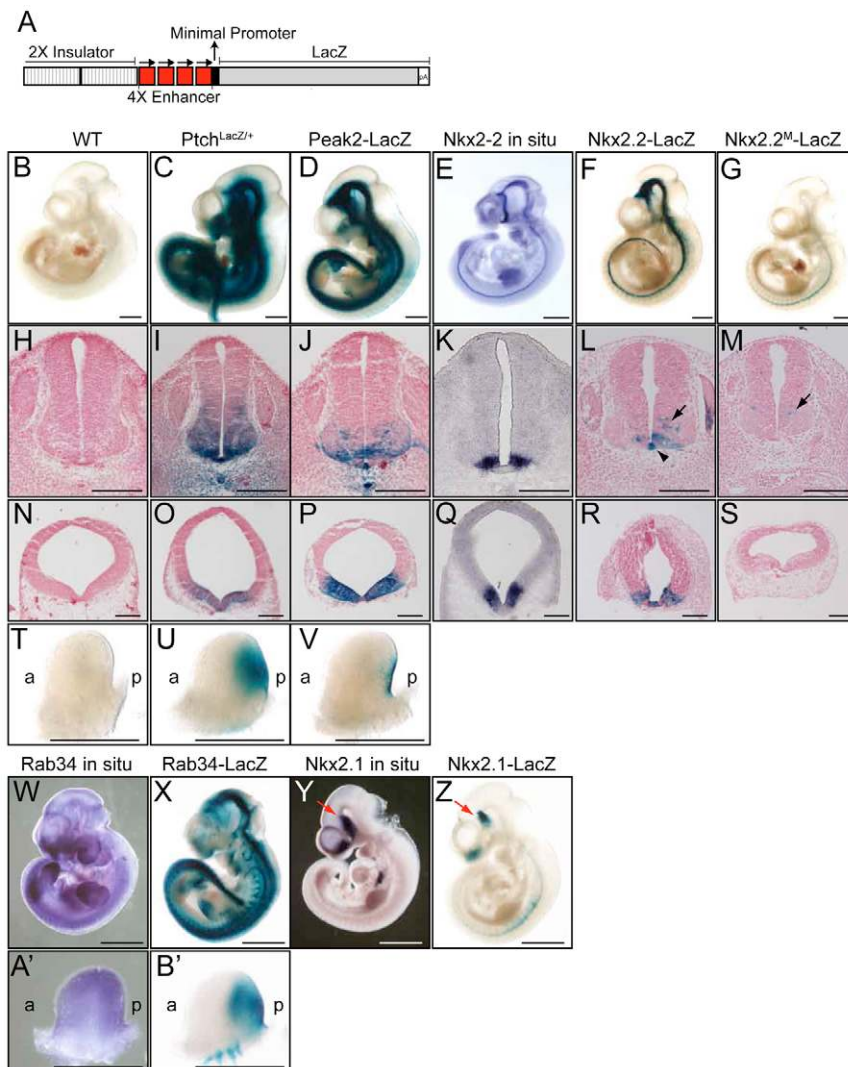
Fig. 3. Transgenic validation of Gli-binding regions demonstrates Gli-dependent expression in Hh target tissue. Selected candidate enhancers (*Ptch1* peak2 and *Nkx2.2* peak1, *Rab34* peak1 and *Nkx2.1* peak1) were cloned into minimal promoter *lacZ* reporter construct (**A**) to generate transgenic embryos and assayed at embryonic day 10.5. (**B-G,W-Z**) The embryo in panel E is cleared in benzyl alcohol and benzyl benzoate; all other whole mounts are in 80% glycerol. Histological sections through the spinal cord (forelimb level) (**H-M**) and brain (**N-S**) of transgenic and control embryos. Forelimbs (**T-V,A′,B′**) show are dissected from the corresponding whole mount. Negative control embryos and sections are shown in B,H,N,T. When compared with *Ptch^lacZ/+^* embryos, *Ptch1* peak2-*LacZ* transgenics express a subset of the normal domain β-gal activity (compare C,I,O,U with D,J,P,V), lacking expression in the limb bud mesenchyme (U,V). In situ hybridizations of *Nkx2.2* (E,K,Q), transgenic Nkx2.2-*lacZ* embryos (F,L,R) or transgenics with the Gli-binding site mutated (G,M,S). In situ hybridizations of *Rab34* (W,A′) and transgenic Rab34-*lacZ* embryos (X,B′). In situ hybridizations of *Nkx2.1* (Y) and transgenic Nkx2.1-*lacZ* embryo (Z). The arrows in X and Z indicate the domain of expression within the ventral diencephalon. Unlike the other transgenics, Nkx2.1-*lacZ* is driven by only one copy of the enhancer. All limb specimens are oriented with anterior to the left and distal up. Scale bars: 200 μm in H-S; 1 mm in A-G,T-Z,A′,B′.

enhancer region (Fig. 2A; Table 2) representing a global target gene. As tissue-specific candidates, we chose a 239 bp region (a region non-responsive in the NIH3T3 assay) representing the single, Gli target prediction within a peak located approximately 1.9 kb 5′ of the TSS in Nkx2.2, a transcriptional regulator with neurally restricted activity associated with V3 interneuron progenitor specification (Fig. 2B; Table 2). Very recently, this site was shown to be Gli responsive (Lei et al., 2006). In addition, we selected a 575 bp region encompassing the single peak in Nkx2.1 11.5 kb downstream of the TES (Fig. 2F) of this forebrain restricted gene (see Fig. S3F in the supplementary material). Finally, we selected a 291 bp region around the single, novel peak in Rab34 intron 1 (see above, Fig. 2D).

Three of the four independent *Ptch1* Peak2-*lacZ* transgenic founder lines exhibited similar β-gal activity. We focused on one of these lines for more extensive analysis, comparing the activity with that of *Ptch1^lacZ/+^* reporter embryos in which a *lacZ* cassette had been knocked into the *Ptch1* locus (Goodrich et al., 1997). *Ptch1* Peak2-*lacZ* transgenics expressed β-gal in the ventral forebrain, midbrain, hindbrain, spinal cord and in the notochord in a similar, although more ventrally restricted and mosaic, pattern within the developing CNS compared with *Ptch1^lacZ/+^* embryos (Fig. 3C,I,O with D,J,P). However, the absence of reporter expression in the branchial arches and most notably in the limb mesenchyme indicates

that other cis-regulatory regions are essential for these components of the 'generic' response to Shh signaling (Fig. 3U,V). Several expected aspects of Hh regulation were observed at later stages in Shh and Ihh target fields (e.g. whisker and chondrocytes respectively, data not shown). In summary, *Ptch1* peak 2 is sufficient for the correct spatial and temporal readout of Hh signaling in some but not all Hh target tissues. Further, its activity in the neural tube is consistent with its identification in the neuralized EB patterning assay.

Nkx2.2-lacZ transgenic embryos were analyzed directly at E10.5; nine of 15 showed detectable β-gal activity; of these, seven had ventral-specific staining in the presumptive brain and spinal cord that clearly encompassed the normal *Nkx2.2* domain (Fig. 3E,F,K,L,Q,R). Examination of sections indicated two ectopic regions of transgenic activity when compared to *Nkx2.2* expression in the neural tube at this stage (Fig. 3E,K,Q). One was in the floor plate (arrowhead in Fig. 3L), the other dorsal to the normal *Nkx2.2* domain (arrow in Fig. 3L). To determine whether the observed expression was Gli-dependent, the single predicted Gli site was mutated, and expression of the resulting Nkx2.2^M^ *lacZ* transgene was analyzed in G0 embryos at E10.5. In contrast to the previous results, only four out of 11 transgenic embryos showed any β-gal activity, and activity was generally weaker and more restricted. Expression in the normal Nkx2.2 domain (in both the brain and

spinal cord) and ectopic expression in the floor plate were entirely dependent upon the Gli-binding region (Fig. 3G,M,S); the weak ectopic, ventrolateral spinal cord domain, however, was Gli-site-independent (arrowhead in Fig. 3M). In summary, this peak region encodes a CRM that regulates Gli-dependent expression of *Nkx2.2* in a domain that encompasses its native expression domain. However, Gli-site dependent expression in the FP region indicates that additional cis-regulatory elements are required to repress *Nkx2.2* in the floor plate; *Nkx2.2* is first activated within ventral midline cells and is co-expressed together with the FP determinant FoxA2 at the outset of FP formation (Jeong and McMahon, 2005). In addition, ectopic ventrolateral Gli-independent activity suggests that additional regulatory regions normally silence non-Hh dependent activation of Nkx2.2 outside its normal domain. Previous data indicate that *Pax6*, and potentially, *Tcf4* play an important role in the ventral restriction of Nkx2.2 (Ericson et al., 1997; Lei et al., 2006).

Nkx2.1 exhibits neural, Hh-dependent expression in a domain in the ventral diencephalon and an additional domain in the ventral telencephalon. Of the two transgenics we obtained, one clearly showed β-gal activity specifically within the expected ventral diencephalic domain (arrows in Fig. 2Y,Z). The absence of expression within the ventral telencephalon suggests that additional sequences are required for expression in this region. Given that the in vitro assay system reflects more posterior neural progenitors, the more anterior progenitors representing the telencephalon were not expected.

Rab34 has a broad expression pattern that would not normally be suggestive of a Hh target gene (Fig. 3W). Nonetheless, transgenic embryos expressed a restricted subset of this expression that correlates well with Hh-responsive regions – 11/17 transgenics expressed β-gal and 10/11 were in a similar, restricted pattern (Fig. 3X). For example, expression of *Rab34* in the limb bud is diffuse, but β-gal driven by the predicted Gli target region was sharply restricted to the posterior, Hh-responsive portion of the limb bud in 7/10 transgenics (Fig. 3A′,B′). This expression seemed to commence after the onset of Hh expression, as the three transgenics without limb bud expression were slightly younger and only the oldest transgenics exhibited expression within the later arising hindlimb.

Gli target specificity is determined by the nuclear availability of Gli1

After characterizing these enhancers, we asked if Gli1 could bind to targets in the absence of gene expression. As shown in Fig. 1, Nkx2.2, FoxA2 and other markers of Hh-dependent ventral neuronal cells were not expressed in RA-treated EBs in the absence of Hh agonist. We examined the degree of Gli1 binding to the enhancer sites of several of these tissue-restricted genes as well as the degree of Gli1 binding to several enhancers that have more global response using the standard ChIP assay. As an additional experiment, we also asked whether we could facilitate Gli1 binding to targets by sequestering it in the nucleus by treatment with leptomycin B immediately prior to harvesting.

Our experiments demonstrate that tissue-specific Gli enhancers generally contain little or no Gli1 binding in RA-only treated EBs. For example, the Nkx2.2 enhancer exhibited 6.9-fold enrichment in ChIPs from Hh-stimulated EBs where gene expression occurs, but was enriched only 1.3-fold in RA-only samples; the Nkx2.9 enhancer was enriched 10-fold in Hh-stimulated versus 1.7-fold in the RA-only samples. Similar trends were seen for Nkx2.1 and FoxA2 (data not shown). In contrast, global enhancers were

Table 3. Predicted Gli target genes in the ventral neural tube

Rank	EEL predictions		MCA predictions	
	EEL	qPCR	MCA	qPCR
1	*Ptch1*	+	Ncam1	–
2	BC017647	–	*Ptch1*	+
3	Kif5a	–	Ctnna2	–
4	Ntn1	–	Gpc3	+
5	Ngfr	–	Robo2	+
6	4831426I19Rik	+	*Nkx2-9*	+
7	Abcb8	–	D230005D02Rik	–
8	Olfm1	–	Ebf3	–
9	*Nkx2-9*	+	Ebf1	+
10	2600003E23Rik	–	Ascl1	–
11	Phox2a	–	*FoxA2*	+
12	Rad541	–	3732412D22Rik	NA
13	Hoxb5	–	Lrrtm1	NA
14	Snx26	–	Ssh3	NA
15	1810059G22Rik	–	Hoxb8	–
16	*FoxA2*	+	Gpm6b	–
17	Mtap1b	–	Trpm3	–
18	Cmkor1	NA	Neurog2	–
19	Tgfbr3	NA	Ret	+
20	5730411O18Rik	NA	Pappa	NA
	Total positive	4/17	Total positive	7/16

Predictions, derived using a pool of Hh-upregulated candidates obtained from transcriptional profiling of embryoid bodies, were ranked according to either EEL or MCA and candidate enhancers were tested by qPCR. Underlined genes were also detected in the preceding ChIP-chip studies. Columns marked NA indicate that the cis-regulatory modules were not tested for that gene. EEL, enhancer element locater; MCA, module cluster analysis; qPCR, quantitative real-time polymerase chain reaction.

enriched by Gli1 in both the presence and absence of Hh stimulation. Here, Ptch1 peak 2 was enriched in both RA-only samples (14.5-fold) and in Hh-stimulated EBs (25.2-fold). Enrichment of Gli binding in the absence of Hh stimulation was also seen for the other global enhancers Ptch1 peak1, Ptch2peak2, Rab34 and Gli1 (data not shown). Thus, Gli1 is able to bind to global enhancers in the absence of Hh stimulation, but is unable to bind efficiently to tissue-specific enhancers in the absence of corresponding gene expression. Surprisingly, when RA-only EBs were treated with 5 nM leptomycin B (Lepto) for 4 hours before ChIP to cause nuclear accumulation of Gli1 (see Fig. S1C in the supplementary material), they exhibited marked increases in the levels of binding to tissue-specific enhancers. Gli1 enrichment of the Nkx2.2 enhancers was 3.3-fold in RA+Lepto samples compared with 1.3-fold in RA-only samples; similarly, the Nkx2.9 enhancer was enriched 9.4-fold in RA+Lepto-treated samples compared with 1.7 fold in RA-only samples. Leptomycin treatment caused similar increases in enrichment for Nkx2.1 and FoxA2, while maintaining the levels of enrichment seen previously for global enhancers in RA-only samples (data not shown). These data argue that Gli1 nuclear concentration is a major determinant of enhancer-binding specificity.

In silico prediction of Gli targets in the ventral neural tube

Next we sought to incorporate information from the Gli target identification to develop a general, testable method for predicting Gli targets in a tissue-specific fashion. We generated a pool of candidate targets by transcriptional profiling of EBs in the presence and absence of Hh signaling and selected 249 genes upregulated in response to Hh-Ag treatment, a group that includes relevant Gli targets in the ventral neural tube (see Table S5 in the supplementary material).

DEVELOPMENT

We then applied two distinct methods: one a recently published algorithm, Enhancer Element Locater (EEL) (Hallikas et al., 2006); and the other a novel method, Module Cluster Analysis (MCA). Our ChIP-chip data indicated that several Gli targets contain multiple peaks that presumably correspond to multiple CRMs (e.g. *Ptch1*, *Nkx2.9*). In MCA, Gli-binding affinities from multiple CRMs were combined, and the combined signal was used as a predictor of Gli target genes (see Materials and methods).

When EEL was applied to the 249 genes upregulated in the Hh-treated EBs, the top 20 predicted enhancers, ranked by their combined Gli-binding strength, included five regions identified in our ChIP screens. We performed qPCR on the remaining 15 predictions and verified one additional peak corresponding to an EST (see Table S6 in the supplementary material), resulting in an effective prediction rate of 1/15 enhancer sites.

When MCA was applied on a genome-wide basis, *Ptch1* and *Hhip1* were both present in the top 10 genes (see Table S7 in the supplementary material), suggesting that the combined Gli-binding strength serves as a good predictor of Gli target genes. We intersected this genome-wide dataset with the Hh-EB upregulated gene list to predict Gli target genes specifically within the ventral neural tube. We then ranked all predicted CRMs that were associated with the top 20 neural Gli target genes and tested the top 20 individual CRMs by qPCR on ChIPed material. Within these top-ranking CRMs, 45% (9/20) were indeed Gli-activator targets by this assay (see Table S8 in the supplementary material). This set included six peaks that had been previously identified by our ChIP-chip studies, giving a new discovery rate of 3/14, compared to 1/15 with EEL.

To summarize, 5/22 of the MCA predicted enhancers (or 11/28 including predictions covered by ChIP peaks) were bound by Gli1. At the gene level, at least 7/20 MCA predicted genes were verified to be direct Gli targets. Thus, pooling information from multiple potential CRMs in the MCA approach improved the ability to predict Gli target genes computationally.

DISCUSSION

In this study, we have developed a general approach for detecting targets of Gli transcriptional regulation by ChIP and applied this to a model of Shh-mediated patterning of the developing mammalian neural tube. The approach was validated by the identification of all previously reported ventral neural tube and general Gli-responsive enhancers that were represented on the custom oligonucleotide arrays, and confirmed by multiple statistical and biological methods. These analyses provide considerable confidence that the nine previously uncharacterized peaks present in the top tier of Table 1 represent genuine Gli targets. In total, the six new Gli target genes represented by the peaks in the top tier of Table 1 and the five new target genes predicted by the MCA in silico analysis and validated by qPCR of ChIP products almost double the total number of known mammalian Gli targets (summarized in Table S9 in the supplementary material). Because Hh-mediated neuronal specification is a relatively well-characterized system, this model provides a particularly good opportunity for validating candidate targets. The core of the strategy, a genetically inducible system for epitope-tagging transcription factors, should be broadly applicable to the study of other transcription factors. Further, the method can be adapted to enable Gli target identification in embryonic and adult tissues using either pre-selected or whole-genome tiling arrays. Modifying the endogenous Gli1 transcript to encode an epitope-tagged Gli1 protein is an alternative strategy; however, this approach restricts analysis to the normal Gli1 expression domain, often representing only a subset of the potential target field.

Identification of enhancer elements during neural tube development

Genome-scale transcription factor binding approaches have been used with great success in yeast, and more recently in ES cells (Lee et al., 2006) and *Drosophila* embryos (Alekseyenko et al., 2006; Sandmann et al., 2006). Embryonic analysis in mammalian systems presents several challenges, including the limited number of cells present in a given target population at a crucial regulatory stage and the dynamic expression of transcription factors during development. By modifying existing protocols we were able to generate robust results with an approximately 50-fold reduction in input ChIP product (approximately 2×10^6 cells per ChIP). The degree of enrichment we obtained with the Gli1[FLAG] construct is impressive considering that: (1) Gli1 is predominately cytoplasmic, and Hh signal-directed nuclear accumulation is transient (Kogerman et al., 1999); (2) the epitope-tagged construct must compete with endogenously transcribed, untagged Gli-activator; and (3) the target population is a mosaic in which relatively few cells represent a given ventral progenitor type (Fig. 1Q). Indeed, we were able to identify bona-fide Gli-activator target sites within enhancers in *Nkx2.2* and *FoxA2*, where Nkx2.2[+] and FoxA2[+] cells represented only 10% of the input population. Interestingly, a subset of predicted peaks did not contain any Gli consensus sites. These regions might be directly Gli-responsive despite the absence of a consensus Gli DNA-binding site. Alternatively, they might represent binding sites for other transcriptional regulators that could then make secondary protein-protein interactions with Gli1. While no specific interactions have been demonstrated for Gli-activator forms, Gli3 proteins have previously been shown to interact with members of the histone deacetylase complex (Cheng and Bishop, 2002; Dai et al., 2002) as well as with Smad proteins (Liu et al., 1998).

In addition to previously characterized *FoxA2* and *Nkx2.9* enhancers, we identified *Nkx2.2*, *Nkx2.1* and *Rab34* as direct targets through transgenic analysis and several new additional targets through validated bioinformatics predictions. However, direct transgenic analysis of *Nkx2.2* illustrates one major challenge the researcher faces in constructing regulatory circuitry from this data: that is, the identification of other cis-regulatory regions that act in conjunction with Gli factors to give appropriate temporal and spatial regulation. In the example of the *Nkx2.2* element we assayed, sequences that repress Gli-dependent FP expression and silence non-Gli dependent ventrolateral neural expression were not present and remain to be identified.

Surprisingly, our data indicate that a global Hh response represented by *Ptch1*, the best characterized transcriptional response to Hh signaling, may in fact be controlled by cell-type-specific response elements. Our analysis of one of four strong, independent Gli-binding regions demonstrates that the region selected represents a subset of the full range and activity of the *Ptch1* transcriptional response. This leaves open the possibility that other regions may encode limb response elements, or that a specific limb response element that normally acts in conjunction with the Gli site in *Ptch1* peak 2 lies outside the regulatory module assayed here. Our assay for putative enhancers using NIH3T3 cells allows us to predict several additional enhancers that are strong candidates for global regulators of a Hh signaling response (Table 2). Consistent with this interpretation, several genes associated with these enhancers are expressed in multiple Hh-responsive tissues (see Fig. S3 in the supplementary material), and all these targets seem to participate in negative feedback loops to regulate expression levels (see model in Fig. 4A).
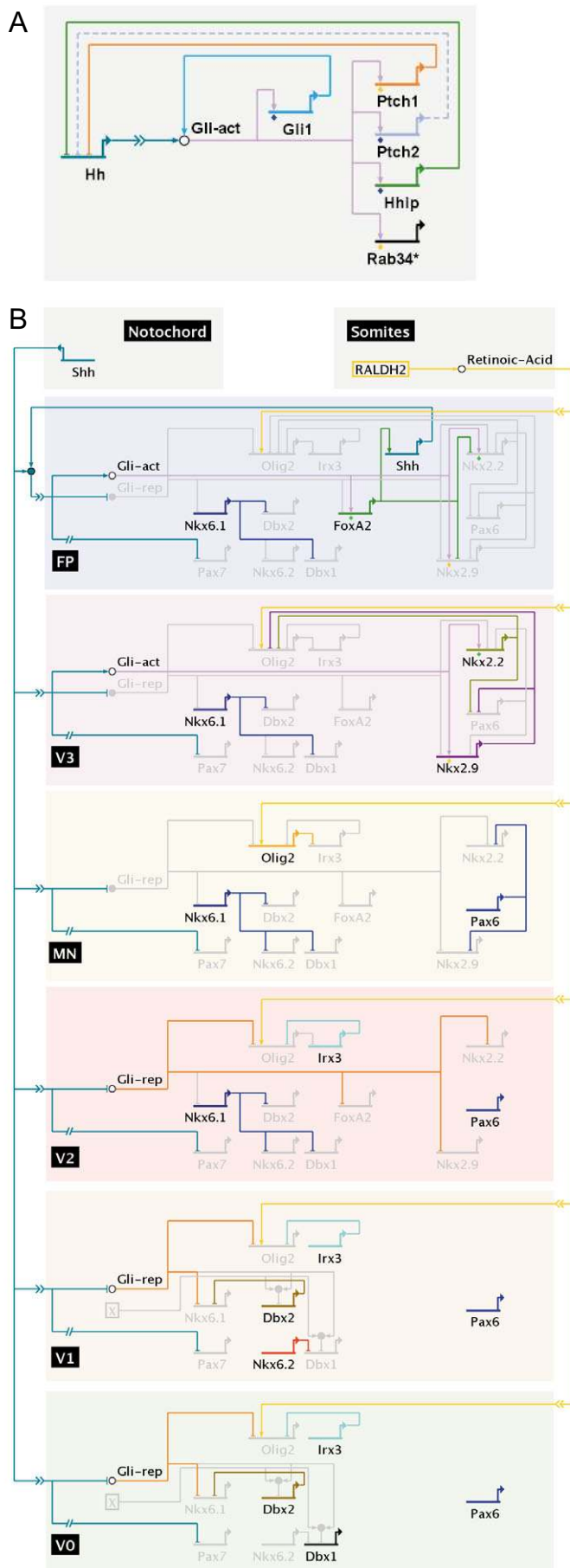
Fig. 4. BioTapestry models of Hh-driven cis-regulatory networks in mouse. Transcriptional targets of the Hh pathway can be defined by a globally responsive signaling cassette (**A**), containing components that are either known (unbroken lines) or likely (broken lines) to be part of a negative feedback loop. (**B**) A model for a Shh-driven transcriptional network underlying ventral neuronal specification. In this diagram, depicted in standard BioTapestry nomenclature (Longabaugh et al., 2005), neuronal specification is depicted as a sequential series of cell states. All genes not expressed are in gray, whereas currently expressed genes are in black or other colors. Similarly, inactive links (active in previous stages of specification) are depicted in gray, whereas active activation or repression is depicted using lines of other colors. An animation of these events can be viewed in Movie 1 in the supplementary material. This diagram focuses explicitly on ventral cell specification; thus previous events in general neuronal specification are not shown. Validated Gli targets (all identified or confirmed in our study) are indicated by blue diamonds (ChIP peak), orange diamonds (transgenic validation) or green diamonds (mutation of binding site in transgenic embryos).

Among the novel targets, *Rab34* is particularly intriguing given the role of another small GTP-binding protein, Rab23, as a crucial regulator of Hh responsiveness (Eggenschwiler et al., 2006; Eggenschwiler et al., 2001) and an independent link between Smo and Ptch1 activity in endosomal trafficking (Zhu et al., 2003). While transcriptional profiling identified *Rab34* as a general, positive target of Hh action in many tissues, *Rab34* is itself rather broadly expressed and therefore, based on expression alone, would not be viewed as a strong candidate for a specific role in the Hh pathway. However, the ChIP data, coupled with studies of Rab34 expression in *Ptch1* and Smo mutants (see Fig. S3D-D″ in the supplementary material), indicate that a component of the Rab34 expression pattern is Hh/Gli-dependent, an observation supported by our transgenic studies (Fig. 2W,X,A′,B′).

Gli sites and target gene regulation

Our data have allowed us to develop, and importantly, to then experimentally test, improved algorithms for the detection of Gli-binding sites, a significant advance for an experimental mammalian dataset. By combining ChIP analysis with transcriptional profiling data and MCA analysis, we were able to correctly predict Gli target genes with an optimal predictive rate of greater than 40% accuracy. Surprisingly, MCA predicts multiple Gli sites within a given locus, whereas only a subset were actually bound by Gli1. This raises the question of whether these are bound by other Gli factors, and if so, whether there is a mechanism that could impart specificity given that all Gli factors bind a similar consensus sequence (Hallikas et al., 2006; Vortkamp et al., 1995). One possibility is that in vivo target sites may have differential affinities for Gli factors.

As noted, we have failed to detect Gli-binding sites in 150 kb of tiled sequence flanking the *Nkx6.1*, one of the very earliest responders to the Hh pathway (see Table S1 in the supplementary material) (Jeong and McMahon, 2005). Whereas the possibility of a more distant Gli1-binding enhancer region remains open, an alternative explanation is that Gli activator binds only to the subset of Gli targets that are expressed in the ventralmost FP and pV3 domains. Because Nkx6.1 is still expressed in mice lacking the activity of all Gli proteins (both activator and repressor) in the neural tube, its expression depends principally upon loss of Gli repression rather than Gli activation (Bai et al., 2004; Lei et al., 2004; Wijgerde et al., 2002).

Our findings, together with published reports that have documented the relationship between Gli factors and Shh signaling in specification and the interactions among specific transcriptional regulators downstream of Shh input, lead to the model depicted in Fig. 4B (Briscoe et al., 2000; Ericson et al., 1997; Fogarty et al., 2005; Novitch et al., 2003; Santagati et al., 2003; Sasaki and Hogan, 1996; Vallstedt et al., 2001; Wijgerde et al., 2002). Shh signaling within the neural tube is effectively interpreted by region-specific combinations of Gli activator and Gli repressor forms. Initially only two cell types are present – one is a V0/V1 bipotential progenitor that depends only indirectly on a loss of Gli repression, which would in turn result in the loss of an inhibitory dorsal signal (Fogarty et al., 2005; Wijgerde et al., 2002). In the second cell state, there is an initial V2 progenitor, which would then sequentially give rise to V2-FP progenitor domains. Most dorsally, where Shh is never received, Gli repressors silence all Hh target genes. At the dorsal limits of Hh signaling (at the dorsal-ventral intersect), Gli repressor would be reduced, thereby relieving the inhibitory action of an unknown factor X on Dbx1 and Dbx2. Dorsal BMP signals have previously been hypothesized to play such a role (Wijgerde et al., 2002); this interaction with Gli repressor is especially attractive, as Gli3 has previously been shown to bind Smad proteins, transcriptional mediators of BMP signaling (Liu et al., 1998). Importantly, the regulation of these domains would not depend on any differential activity of Gli repressor – only on the attenuation of the dorsal inhibitory input. In domains that receive a higher Hh input, Gli repression would be progressively attenuated to the pMN domain, and loss of this repressor is sufficient for activation of targets. Within the pV3 and FP domains, however, specification requires varying levels of Gli-activator activity. Recent experiments in chick argue persuasively for a ratiometric assessment of Gli repressor and activator in dictating the position-specific Hh response both in the limb and the neural tube (Davey et al., 2006) (E. Dessaud and J. Briscoe, personal communication). Finally, the available data indicate an important dynamic component to ventral patterning, in which specific neural progenitor populations may change their identity on integrating Hh signaling over time (see Movie 1 in the supplementary material).

In summary, by applying biochemical and bioinformatics approaches, we have a new view of Hh targets, their regulation and the general parameters that characterize a Gli-activator response in patterning the vertebrate neural tube. These data provide a solid grounding for explaining the possible roles of newly defined targets (e.g. Rab34) and for determining the relationships between Gli proteins and other transcriptional regulators in coordinating precise transcriptional outcomes through local cis-acting regulatory modules in central Hh target genes.

Supplementary material

Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/134/10/1977/DC1

References

Agren, M., Kogerman, P., Kleman, M. I., Wessling, M. and Toftgard, R. (2004). Expression of the PTCH1 tumor suppressor gene is regulated by alternative promoters and a single functional Gli-binding site. *Gene* **330**, 101-114.

Alekseyenko, A. A., Larschan, E., Lai, W. R., Park, P. J. and Kuroda, M. I. (2006). High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev.* **20**, 848-857.

Bai, C. B. and Joyner, A. L. (2001). Gli1 can rescue the in vivo function of Gli2. *Development* **128**, 5161-5172.

Bai, C. B., Stephen, D. and Joyner, A. L. (2004). All mouse ventral spinal cord patterning by hedgehog is Gli dependent and involves an activator function of Gli3. *Dev. Cell* **6**, 103-115.

Briscoe, J. and Ericson, J. (2001). Specification of neuronal fates in the ventral neural tube. *Curr. Opin. Neurobiol.* **11**, 43-49.

Briscoe, J., Sussel, L., Serup, P., Hartigan-O'Connor, D., Jessell, T. M., Rubenstein, J. L. and Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. *Nature* **398**, 622-627.

Briscoe, J., Pierani, A., Jessell, T. M. and Ericson, J. (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* **101**, 435-445.

Cheng, S. Y. and Bishop, J. M. (2002). Suppressor of Fused represses Gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex. *Proc. Natl. Acad. Sci. USA* **99**, 5442-5447.

Chuang, P. T. and McMahon, A. P. (1999). Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* **397**, 617-621.

Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190.

Dai, P., Akimaru, H., Tanaka, Y., Maekawa, T., Nakafuku, M. and Ishii, S. (1999). Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. *J. Biol. Chem.* **274**, 8143-8152.

Dai, P., Shinagawa, T., Nomura, T., Harada, J., Kaul, S. C., Wadhwa, R., Khan, M. M., Akimaru, H., Sasaki, H., Colmenares, C. et al. (2002). Ski is involved in transcriptional regulation by the repressor and full-length forms of Gli3. *Genes Dev.* **16**, 2843-2848.

Davey, M. G., Paton, I. R., Yin, Y., Schmidt, M., Bangs, F. K., Morrice, D. R., Smith, T. G., Buxton, P., Stamataki, D., Tanaka, M. et al. (2006). The chicken talpid3 gene encodes a novel protein essential for Hedgehog signaling. *Genes Dev.* **20**, 1365-1377.

Davidson, E. H. and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800.

Eggenschwiler, J. T., Espinoza, E. and Anderson, K. V. (2001). Rab23 is an essential negative regulator of the mouse Sonic hedgehog signalling pathway. *Nature* **412**, 194-198.

Eggenschwiler, J. T., Bulgakov, O. V., Qin, J., Li, T. and Anderson, K. V. (2006). Mouse Rab23 regulates hedgehog signaling from smoothened to Gli proteins. *Dev. Biol.* **290**, 1-12.

Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., van Heyningen, V., Jessell, T. M. and Briscoe, J. (1997). Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell* **90**, 169-180.

Fogarty, M., Richardson, W. D. and Kessaris, N. (2005). A subset of oligodendrocytes generated from radial glia in the dorsal spinal cord. *Development* **132**, 1951-1959.

Frank-Kamenetsky, M., Zhang, X. M., Bottega, S., Guicherit, O., Wichterle, H., Dudek, H., Bumcrot, D., Wang, F. Y., Jones, S., Shulok, J. et al. (2002). Small-molecule modulators of Hedgehog signaling: identification and characterization of Smoothened agonists and antagonists. *J. Biol.* **1**, 10.

Goodrich, L. V., Milenkovic, L., Higgins, K. M. and Scott, M. P. (1997). Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109-1113.

Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E. and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124**, 47-59.

Hooper, J. E. and Scott, M. P. (2005). Communicating with Hedgehogs. *Nat. Rev. Mol. Cell Biol.* **6**, 306-317.

Jacob, J. and Briscoe, J. (2003). Gli proteins and the control of spinal-cord patterning. *EMBO Rep.* **4**, 761-765.

Jeong, J. and McMahon, A. P. (2005). Growth and pattern of the mammalian neural tube are governed by partially overlapping feedback activities of the hedgehog antagonists patched 1 and Hhip1. *Development* **132**, 143-154.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Rev. Genet.* **1**, 20-29.

Ji, H. and Wong, W. H. (2005). TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**, 3629-3636.

Kinzler, K. W. and Vogelstein, B. (1990). The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol. Cell. Biol.* **10**, 634-642.

Kogerman, P., Grimm, T., Kogerman, L., Krause, D., Unden, A. B., Sandstedt, B., Toftgard, R. and Zaphiropoulos, P. G. (1999). Mammalian suppressor-of-fused modulates nuclear-cytoplasmic shuttling of Gli-1. *Nat. Cell Biol.* **1**, 312-319.

Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F. and Wootton, J. C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **262**, 208-214.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K. et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313.

Lei, Q., Zelman, A. K., Kuang, E., Li, S. and Matise, M. P. (2004). Transduction of graded Hedgehog signaling by a combination of Gli2 and Gli3 activator functions in the developing spinal cord. *Development* **131**, 3593-3604.

Lei, Q., Jeong, Y., Misra, K., Li, S., Zelman, A. K., Epstein, D. J. and Matise, M. P. (2006). Wnt signaling inhibitors regulate the transcriptional response to morphogenetic Shh-Gli signaling in the neural tube. *Dev. Cell* **11**, 325-337.

Li, C. and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* **98**, 31-36.

Liu, F., Massague, J. and Ruiz i Altaba, A. (1998). Carboxy-terminally truncated Gli3 proteins associate with Smads. *Nat. Genet.* **20**, 325-326.

Liu, J. S. (1994). The collapsed Gibbs sampler with applications to a gene regulation problem. *J. Am. Stat. Assoc.* **89**, 958-966.

Longabaugh, W. J. R., Davidson, E. H. and Bolouri, H. (2005). Computational representation of developmental genetic regulatory networks. *Dev. Biol.* **283**, 1-16.

Mao, J., Barrow, J., McMahon, J., Vaughan, J. and McMahon, A. P. (2005). An ES cell system for rapid, spatial and temporal analysis of gene function in vitro and in vivo. *Nucleic Acids Res.* **33**, e155.

McMahon, A. P., Ingham, P. W. and Tabin, C. J. (2003). Developmental roles and clinical significance of hedgehog signaling. *Curr. Top. Dev. Biol.* **53**, 1-114.

Megason, S. G. and McMahon, A. P. (2002). A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development* **129**, 2087-2098.

Motoyama, J., Takabatake, T., Takeshima, K. and Hui, C. (1998). Ptch2, a second mouse Patched gene is co-expressed with Sonic hedgehog. *Nat. Genet.* **18**, 104-106.

Muller, B. and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. *Development* **127**, 2999-3007.

Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K. and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87-90.

Novitch, B. G., Wichterle, H., Jessell, T. M. and Sockanathan, S. (2003). A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. *Neuron* **40**, 81-95.

Nybakken, K., Vokes, S. A., Lin, T. Y., McMahon, A. P. and Perrimon, N. (2005). A genome-wide RNA interference screen in Drosophila melanogaster cells for new components of the Hh signaling pathway. *Nat. Genet.* **37**, 1323-1332.

Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K. et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378-1381.

Pabst, O., Herbrand, H., Takuma, N. and Arnold, H. H. (2000). NKX2 gene expression in neuroectoderm but not in mesendodermally derived structures depends on sonic hedgehog in mouse embryos. *Dev. Genes Evol.* **210**, 47-50.

Pan, Y., Bai, C. B., Joyner, A. L. and Wang, B. (2006). Sonic hedgehog signaling regulates Gli2 transcriptional activity by suppressing its processing and degradation. *Mol. Cell. Biol.* **26**, 3365-3377.

Pavletich, N. P. and Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261**, 1701-1707.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-2309.

Ruiz i Altaba, A., Nguyen, V. and Palma, V. (2003). The emergent design of the neural tube: prepattern, SHH morphogen and GLI code. *Curr. Opin. Genet. Dev.* **13**, 513-521.

Sandmann, T., Jensen, L. J., Jakobsen, J. S., Karzynski, M. M., Eichenlaub, M. P., Bork, P. and Furlong, E. E. (2006). A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. *Dev. Cell* **10**, 797-807.

Santagati, F., Abe, K., Schmidt, V., Schmitt-John, T., Suzuki, M., Yamamura, K. and Imai, K. (2003). Identification of Cis-regulatory elements in the mouse Pax9/Nkx2-9 genomic region: implication for evolutionary conserved synteny. *Genetics* **165**, 235-242.

Sasaki, H. and Hogan, B. L. (1996). Enhancer analysis of the mouse HNF-3 beta gene: regulatory elements for node/notochord and floor plate are independent and consist of multiple sub-elements. *Genes Cells* **1**, 59-72.

Sasaki, H., Hui, C., Nakafuku, M. and Kondoh, H. (1997). A binding site for Gli proteins is essential for HNF-3beta floor plate enhancer activity in transgenics and can respond to Shh in vitro. *Development* **124**, 1313-1322.

Srinivas, S., Watanabe, T., Lin, C. S., William, C. M., Tanabe, Y., Jessell, T. M. and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4.

Stamataki, D., Ulloa, F., Tsoni, S. V., Mynett, A. and Briscoe, J. (2005). A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev.* **19**, 626-641.

Teboul, L., Summerbell, D. and Rigby, P. W. (2003). The initial somitic phase of Myf5 expression requires neither Shh signaling nor Gli regulation. *Genes Dev.* **17**, 2870-2874.

Tenzen, T., Allen, B. L., Cole, F., Kang, J. S., Krauss, R. S. and McMahon, A. P. (2006). The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev. Cell* **10**, 647-656.

Vallstedt, A., Muhr, J., Pattyn, A., Pierani, A., Mendelsohn, M., Sander, M., Jessell, T. M. and Ericson, J. (2001). Different levels of repressor activity assign redundant and specific roles to Nkx6 genes in motor neuron and interneuron specification. *Neuron* **31**, 743-755.

Vortkamp, A., Gessler, M. and Grzeschik, K. H. (1995). Identification of optimized target sequences for the GLI3 zinc finger protein. *DNA Cell Biol.* **14**, 629-634.

Wang, B., Fallon, J. F. and Beachy, P. A. (2000). Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* **100**, 423-434.

Wichterle, H., Lieberam, I., Porter, J. A. and Jessell, T. M. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110**, 385-397.

Wijgerde, M., McMahon, J. A., Rule, M. and McMahon, A. P. (2002). A direct requirement for Hedgehog signaling for normal specification of all ventral progenitor domains in the presumptive mammalian spinal cord. *Genes Dev.* **16**, 2849-2864.

Zhang, X. M., Ramalho-Santos, M. and McMahon, A. P. (2001). Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R symmetry by the mouse node. *Cell* **106**, 781-792.

Zhu, A. J., Zheng, L., Suyama, K. and Scott, M. P. (2003). Altered localization of Drosophila Smoothened protein activates Hedgehog signal transduction. *Genes Dev.* **17**, 1240-1252.