

TUTORIALS

Coding Sequence

Sonic Hedgehog (*SHH*)

1. Where did the name for this gene come from (hint: RefSeq track)?

The video game character *Sonic the Hedgehog*

2. What human diseases are caused by mutations in *SHH*? What's their OMIM number?

Holoprosencephaly-3, *Isolated Microphthalmia with Coloboma*, cleft lip and/or palate (Preaxial Polydactyly II is due to a REGULATORY mutation NOT a gene one); OMIM 600725

3. What chemicals interact with *SHH*?

Tretinoin, 4-(diethylamino)benzaldehyde, Acetaminophen, Estradiol, Ethanol, Forskolin, Lipopolysaccharides, Valproic Acid, Vitamin A, cyclopamine

4. What is the PDB code that ModBase predicts its structure for?

1vhh

5. What is the 'Molecular Function' Gene Ontology (GO) term for *SHH*?

patched binding, protein binding, peptidase activity, hydrolase activity, laminin-1 binding

6. To what other genes does the *SHH* mRNA sequence have similarity to?

IHH, DHH

PROBLEMS 7 & 8 ARE HOMEWORK

7. Make a table of the genomic coordinates of *SHH* 'RefSeq genes' exons.

```
#exonStarts  exonEnds
155592677,155598989,155599404,155601686,    155592939,155599251,155599533,155601766,
155592677,155593151,155598989,155599404,155601686,
    155592939,155593361,155599251,155599533,155601766,
155592677,155593151,155598989,155599404,155601686,
    155592939,155593291,155599251,155599533,155601766,
155594933,155598989,155604516, 155596420,155599251,155604967,
```

8. Make a track that shows the regions overlapping between the *SHH* 'RefSeq' gene and human mRNAs.

```
track name="tb_refGene" description="table browser query on refGene" visibility=3 url=
chr7  155592677      155601766      NM_001310462    0      -      155592700
      155599443      0      4      262,262,129,80,0,6312,6727,9009,
chr7  155592677      155601766      NR_132318      0      -      155601766
      155601766      0      5      262,210,262,129,80,0,474,6312,6727,9009,
chr7  155592677      155601766      NR_132319      0      -      155601766
      155601766      0      5      262,140,262,129,80,0,474,6312,6727,9009,
```

chr7	155594933	155604967	NM_000193	0	-	155595593
	155604816	0	3	1487,262,451,	0,4056,9583,	

Noncoding Sequences

FORWARD: TCATTGTAATTCATAGACCATGCTG
REVERSE: AAAGTTAGAAGATCAAACGAAAGAGG

9. What is the genomic region that these primers pick up?

chrX:25,008,172-25,009,245

10. What would be the genomic region in the March 2006 (hg18) assembly?

chrX:24,918,093-24,919,166

11. Is this sequence evolutionarily conserved?

Yes. This is an ultraconserved element. One of the most conserved sequences in the genome.

12. How many repeats are within this sequence and what are they?

One: Low Complexity

13. How many SNPs are in this region and what's their name?

rs76067703

14. What is the minor allele frequency of this SNP/s?

0.5

PROBLEMS 7 & 8 ARE HOMEWORK

15. Does the sequence overlap an EnhancerFinder predicted brain enhancer? What is its score? (You will need to make a custom track from the brain BED file of EnhancerFinder predictions from Data File S1 in Erwin et al. Note: this file is in hg19 coordinates.)

Yes, score =1.47

16. Was this sequence tested for enhancer activity by the VISTA Enhancer Browser project? If so, in which tissues is it active?

Yes. Forebrain.