The top CNV component has the followinfg trio-states. Where '1' indicates a deletion and order is F, M, O.

```
000 001 010 011 100 101 110 111
326   3  14  35  14  32   1  20
```

Methods

Cleft Data Description

- Performed 284 tests. Bonferroni significant locus has width 4.7 kB.
- *PennCNV* joint HMM
- european, MAD < 0.3, non-WGA, aux $\neq 1$
- coverage > 10
- 13140 hemi/homozygous deletions identified in 445 trios
- 4288 CNV components

  Common (> 0.01): 954

  Rare: 3334
- Construct trio-states for all CNV components

  recall that we use indicator variable for hemi/homozygous deletions
- must be at least 5 informative mating pairs

  01x and 10x
- count transmissions and perform binom.test (See "trans.tab")

```
> est.list.beaty <- lapply(binom.list.beaty, FUN = trioClasses:::get.est)
> ci.list.beaty <- lapply(binom.list.beaty, FUN = trioClasses:::get.ci)
> testable.beaty <- !is.na(ci.list.beaty)
> ci.mat.beaty <- matrix(unlist(ci.list.beaty[testable.beaty]),
    nrow = sum(testable.beaty), ncol = 2, byrow = TRUE)
> est.vec.beaty <- as(est.list.beaty, "numeric")[testable.beaty]
> cnv.beaty.obj$cmp.gr[testable.beaty]
```

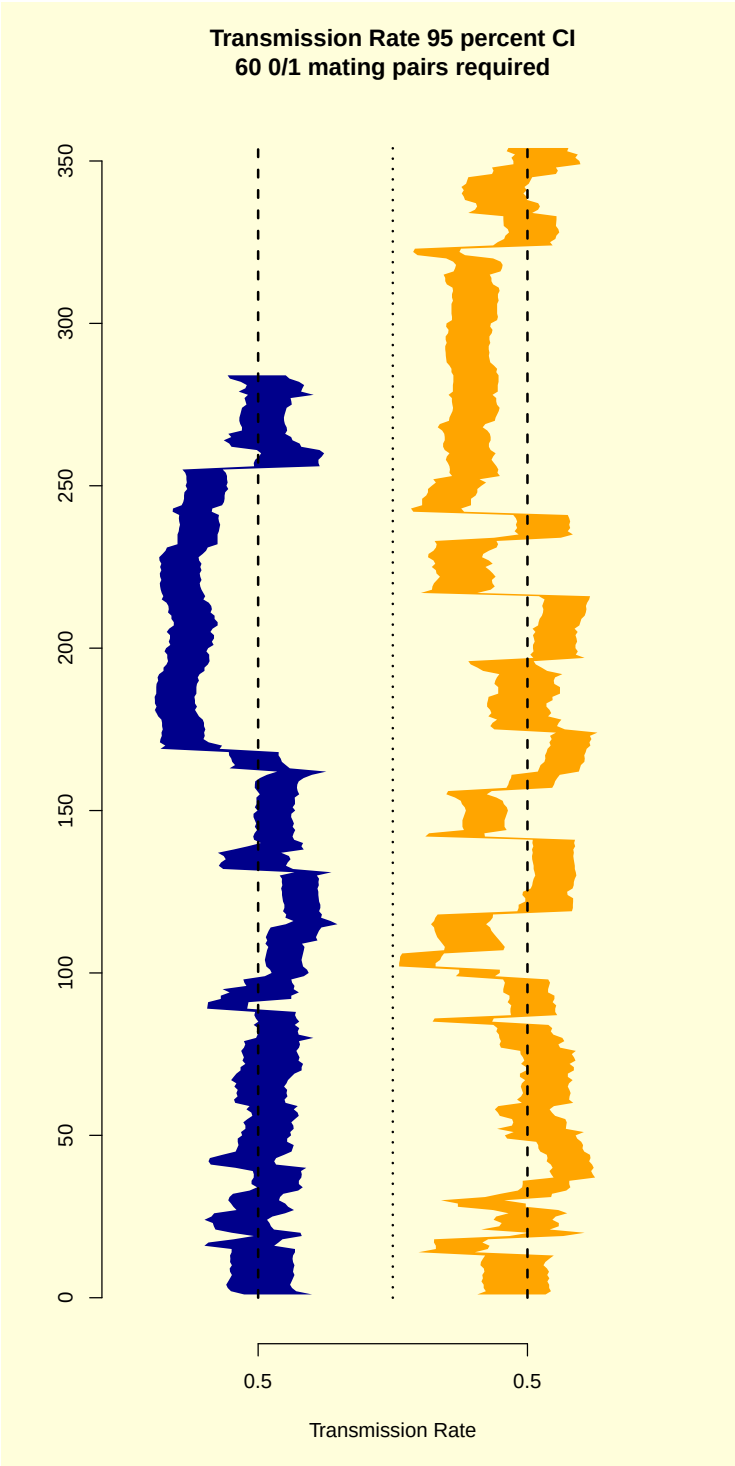
```
GRanges with 284 ranges and 0 metadata columns:
           seqnames                ranges strand
              <Rle>             <IRanges>  <Rle>
   comp120     chr1 [103969301, 103977611]      *
   comp121     chr1 [103977612, 103983731]      *
   comp122     chr1 [103983732, 103984810]      *
   comp123     chr1 [103984811, 103988830]      *
   comp124     chr1 [103988831, 103989217]      *
```

```
      ...      ...                     ...    ...
  comp3942    chr19   [20418019, 20422199]      *
  comp3943    chr19   [20422200, 20493452]      *
  comp3944    chr19   [20493453, 20507068]      *
  comp4116    chr20   [52081230, 52081774]      *
  comp4117    chr20   [52081775, 52088118]      *
  ---
  seqlengths:
            chr1   chr1_random          chr2 ...          chrY          chrM
       247249719       1663265     242951149 ...      57772954         16571
```

```
> est.list.pitt <- lapply(binom.list.pitt, FUN = trioClasses:::get.est)
> ci.list.pitt <- lapply(binom.list.pitt, FUN = trioClasses:::get.ci)
> testable.pitt <- !is.na(ci.list.pitt)
> ci.mat.pitt <- matrix(unlist(ci.list.pitt[testable.pitt]),
     nrow = sum(testable.pitt), ncol = 2, byrow = TRUE)
> est.vec.pitt <- as(est.list.pitt, "numeric")[testable.pitt]
> cnv.pitt.obj$cmp.gr[testable.pitt]
```

```
GRanges with 354 ranges and 0 metadata columns:
           seqnames                ranges strand
              <Rle>             <IRanges>  <Rle>
   comp117     chr1 [103977612, 103983731]      *
   comp118     chr1 [103983732, 103989830]      *
   comp119     chr1 [103989831, 103990016]      *
   comp120     chr1 [103990017, 103991756]      *
   comp121     chr1 [103991757, 103999836]      *
       ...      ...                     ...    ...
  comp3774    chr20   [52081775, 52088118]      *
  comp3831    chr21   [14163409, 14164422]      *
  comp3832    chr21   [14164423, 14166029]      *
  comp3833    chr21   [14166030, 14166821]      *
  comp3834    chr21   [14166822, 14169701]      *
  ---
  seqlengths:
            chr1   chr1_random          chr2 ...          chrY          chrM
       247249719       1663265     242951149 ...      57772954         16571
```

The CNV components with signiifcant p-values (Bonferroni) are given below.


Transmission Rate 95 percent CI
60 0/1 mating pairs required
350
300
250
200
150
100
50
0
0.5
0.5
Transmission Rate

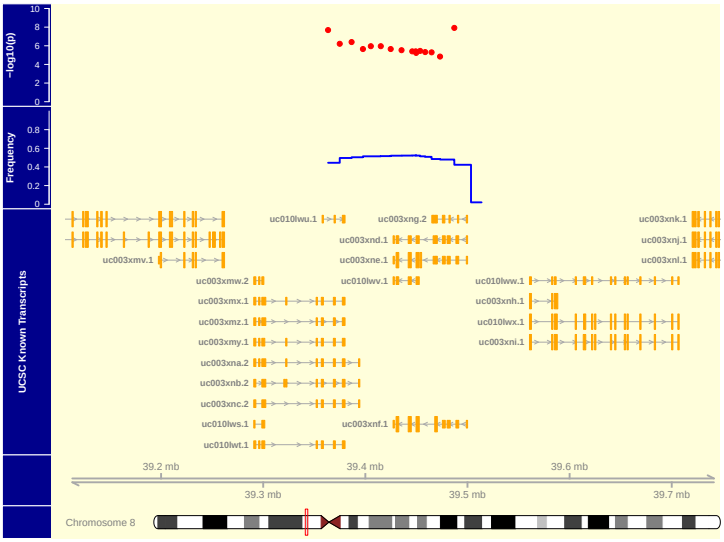


Figure 1: Cleft Trios — Red: $-\log_{10} p$-values for each CNV component that has at least **five** "0/1" mating pairs. Null hypothesis is transmission rate of $\frac{1}{2}$, and alternative hypothesis is "one-sided." Blue: The frequency of CNV component in the parents. Yellow: Gene tracks, labeled by transcript.

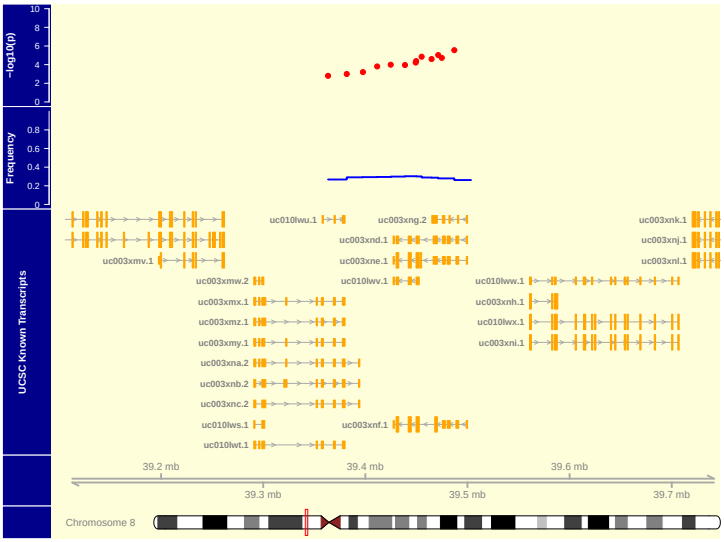


Figure 2: Control Trios — Red: $-\log_{10} p$-values for each CNV component that has at least **five** "0/1" mating pairs. Null hypothesis is transmission rate of $\frac{1}{2}$, and alternative hypothesis is "one-sided." Blue: The frequency of CNV component in the parents. Yellow: Gene tracks, labeled by transcript.

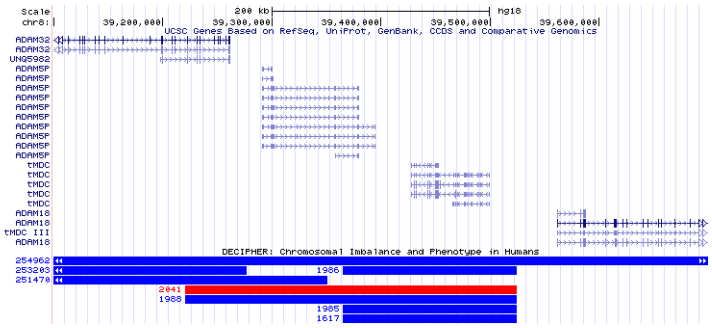


Figure 3: UCSC Genome Browser at chromosome 8 locus.

```
GRanges with 17 ranges and 0 metadata columns:
           seqnames                ranges strand
              <Rle>             <IRanges>  <Rle>
  comp1883     chr7 [141435189, 141439888]      *
  comp1994     chr8 [ 39356825,  39370186]      *
  comp1995     chr8 [ 39370187,  39379683]      *
  comp1996     chr8 [ 39379684,  39393417]      *
  comp1997     chr8 [ 39393418,  39401865]      *
       ...      ...                   ...    ...
  comp2005     chr8   [39450168, 39457081]      *
  comp2006     chr8   [39457082, 39460087]      *
  comp2007     chr8   [39460088, 39469612]      *
  comp2008     chr8   [39469613, 39476658]      *
  comp2009     chr8   [39476659, 39497557]      *
  ---
  seqlengths:
            chr1   chr1_random          chr2 ...          chrY          chrM
       247249719       1663265     242951149 ...      57772954         16571
```

All CNV components are contiguous and the total width is 4.7 kB.

```
function (object)
{
    if (sum(object[c("101", "011", "100", "010")], na.rm = TRUE) >=
        60) {
        T <- sum(object[c("101", "011")], na.rm = TRUE)
        U <- sum(object[c("100", "010")], na.rm = TRUE)
        return(binom.test(x = T, n = T + U, p = 0.5, alternative = "greater")$p.value)
    }
    else {
        return(NA)
    }
}
```

```
<environment: namespace:trioClasses>

> gr.deletion.pitt <- gr.pitt[values(gr.pitt)$numsnp >= 10 &
     values(gr.pitt)$cn %in% 0:1]
> sum(countOverlaps(gr.deletion.pitt, reduce(locus)))

[1] 649
```