Jochum, Michael D.

CG1808

29 April 2018

Canis lupus familiaris Chromosome 2 gene model comparison
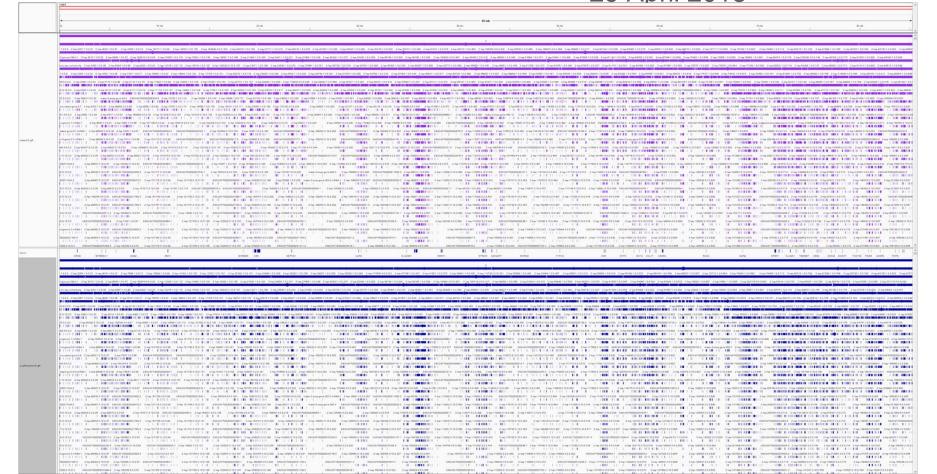
Canis lupus familiaris Chromosome 2 gene model comparison
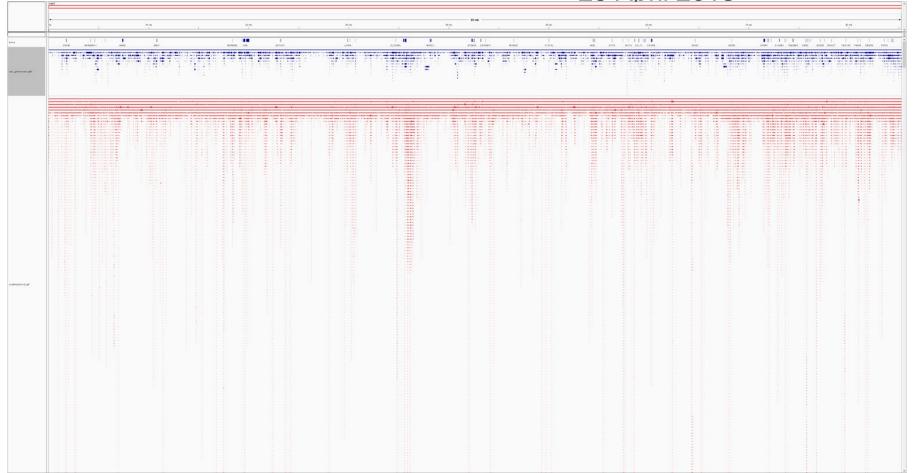
Canis lupus familiaris Chromosome 2 gene model comparison

```
# gffcompare v0.10.1 | Command line was:
#gffcompare -r unaffectedchr2.gff -o fast_vs_unaff fastchr2.gff
#

#= Summary for dataset: fastchr2.gff
#    Query mRNAs :   1034 in   1034 loci  (796 multi-exon transcripts)
#          (0 multi-transcript loci, ~1.0 transcripts per locus)
# Reference mRNAs :   1034 in   1034 loci  (796 multi-exon)
# Super-loci w/ reference transcripts:    1034
#-----------------| Sensitivity | Precision |
        Base level:  100.0   |   100.0   |
        Exon level:  100.0   |   100.0   |
       Intron level:  100.0   |   100.0   |
  Intron chain level:  100.0   |   100.0   |
    Transcript level:  100.0   |   100.0   |
        Locus level:  100.0   |   100.0   |

     Matching intron chains:    796
      Matching transcripts:   1034
            Matching loci:   1034

       Missed exons:     0/6460   (  0.0%)
        Novel exons:     0/6460   (  0.0%)
      Missed introns:    0/5426   (  0.0%)
       Novel introns:    0/5426   (  0.0%)
        Missed loci:     0/1034   (  0.0%)
         Novel loci:     0/1034   (  0.0%)

 Total union super-loci across all input datasets: 1034
1034 out of 1034 consensus transcripts written in
fast_vs_unaff.annotated.gtf (0 discarded as redundant)
```

```
# gffcompare v0.10.1 | Command line was:
#gffcompare -r unaffectedchr2.gff -o slow_vs_unaff slowchr2.gff
#

#= Summary for dataset: slowchr2.gff
#    Query mRNAs :   1034 in   1034 loci  (796 multi-exon transcripts)
#          (0 multi-transcript loci, ~1.0 transcripts per locus)
# Reference mRNAs :   1034 in   1034 loci  (796 multi-exon)
# Super-loci w/ reference transcripts:    1034
#-----------------| Sensitivity | Precision |
        Base level:  100.0   |   100.0   |
        Exon level:  100.0   |   100.0   |
       Intron level:  100.0   |   100.0   |
  Intron chain level:  100.0   |   100.0   |
    Transcript level:  100.0   |   100.0   |
        Locus level:  100.0   |   100.0   |

     Matching intron chains:    796
      Matching transcripts:   1034
            Matching loci:   1034

       Missed exons:     0/6460   (  0.0%)
        Novel exons:     0/6460   (  0.0%)
      Missed introns:    0/5426   (  0.0%)
       Novel introns:    0/5426   (  0.0%)
        Missed loci:     0/1034   (  0.0%)
         Novel loci:     0/1034   (  0.0%)

 Total union super-loci across all input datasets: 1034
1034 out of 1034 consensus transcripts written in slow_vs_unaff.annotated.gtf (0
discarded as redundant)
slow_vs_unaff.stats (END)
```

```
# gffcompare v0.10.1 | Command line was:
#gffcompare -r unaffected_out.gtf -o fast_vs_unaff fast_out.gtf
#

#= Summary for dataset: fast_out.gtf
#    Query mRNAs :  98983 in  76026 loci  (49987 multi-exon
transcripts)
#          (10729 multi-transcript loci, ~1.3 transcripts per locus)
# Reference mRNAs :  53881 in  34894 loci  (45545 multi-exon)
# Super-loci w/ reference transcripts:   30912
#-----------------| Sensitivity | Precision |
        Base level:   87.9   |   55.1   |
        Exon level:   94.2   |   79.5   |
       Intron level:   96.3   |   94.5   |
Intron chain level:   83.0   |   75.6   |
   Transcript level:   83.8   |   45.6   |
        Locus level:   95.1   |   43.6   |

     Matching intron chains:   37787
      Matching transcripts:   45177
            Matching loci:   33176

       Missed exons:    3510/265778  (  1.3%)
        Novel exons:   47226/316131  ( 14.9%)
       Missed introns:    4102/210726  (  1.9%)
        Novel introns:    2425/214850  (  1.1%)
        Missed loci:     519/34894  (  1.5%)
         Novel loci:   40748/76026  ( 53.6%)

 Total union super-loci across all input datasets: 75130
98983 out of 98983 consensus transcripts written in
fast_vs_unaff.annotated.gtf (0 discarded as redundant)
```

```
# gffcompare v0.10.1 | Command line was:
#gffcompare -r ../unaffected_out.gtf -o slow_vs_unaffected ./slow_out.gtf
#= Summary for dataset: ./slow_out.gtf
#    Query mRNAs :  87236 in  65097 loci  (49198 multi-exon transcripts)
#          (10447 multi-transcript loci, ~1.3 transcripts per locus)
# Reference mRNAs :  53881 in  34894 loci  (45545 multi-exon)
# Super-loci w/ reference transcripts:   30980
#-----------------| Sensitivity | Precision |
        Base level:   88.5   |   57.4   |
        Exon level:   94.2   |   82.8   |
       Intron level:   96.2   |   94.8   |
Intron chain level:   82.9   |   76.7   |
   Transcript level:   83.7   |   51.7   |
        Locus level:   95.1   |   50.9   |

     Matching intron chains:   37735
      Matching transcripts:   45119
            Matching loci:   33185

       Missed exons:    3503/265778  (  1.3%)
        Novel exons:   35920/303457  ( 11.8%)
       Missed introns:    4131/210726  (  2.0%)
        Novel introns:    2163/213883  (  1.0%)
        Missed loci:     501/34894  (  1.4%)
         Novel loci:   29869/65097  ( 45.9%)

 Total union super-loci across all input datasets: 64256
87236 out of 87236 consensus transcripts written in
slow_vs_unaffected.annotated.gtf (0 discarded as redundant)
```

	maker fast	maker slow	stringtie fast	stringtie slow
Query mRNAs	1034/1034	1034/1034	98983/76026	87236/65097
reference mRNAs	1034/1034	1034/1034	53881/34894	53882/34894
Base level	100	100	87.9	88.5
Exon level	100	100	94.2	94.2
Intron level	100	100	96.3	96.2
Inton chain level	100	100	83.0	82.9
Transcript level	100	100	83.8	83.7
Locus level	100	100	95.1	95.1
Matching intron chains	796	796	37787	37735
Mathching transcipts	1034	1034	45177	45119
Matching loci	1034	1034	33176	33185
Missed exons	0.0%	0.0%	1.30%	1.30%
Novel exons	0.0%	0.0%	14.90%	11.80%
Missed introns	0.0%	0.0%	1.90%	2.00%
Novel introns	0.0%	0.0%	1.10%	1.00%
Missed loci	0.0%	0.0%	1.50%	1.40%
Novel loci	0.0%	0.0%	53.60%	45.90%
Total union super-loci	1034	1034	75130	64256
consensus transcripts	1034	1034	98983/98983	87236/87236

Author: Michael D Jochum Jr <jochum00@login3>