Genome Assembly using NGS Data Algorithms and Applications

Zemin Ning

The Wellcome Trust Sanger Institute

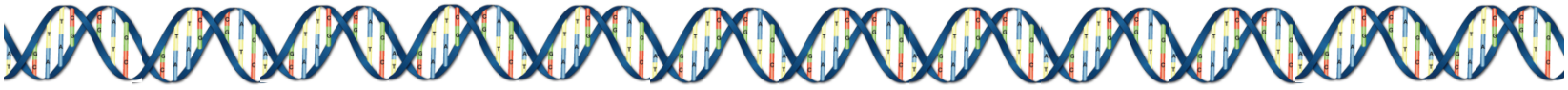
wellcome trust sanger institute

Solexa™

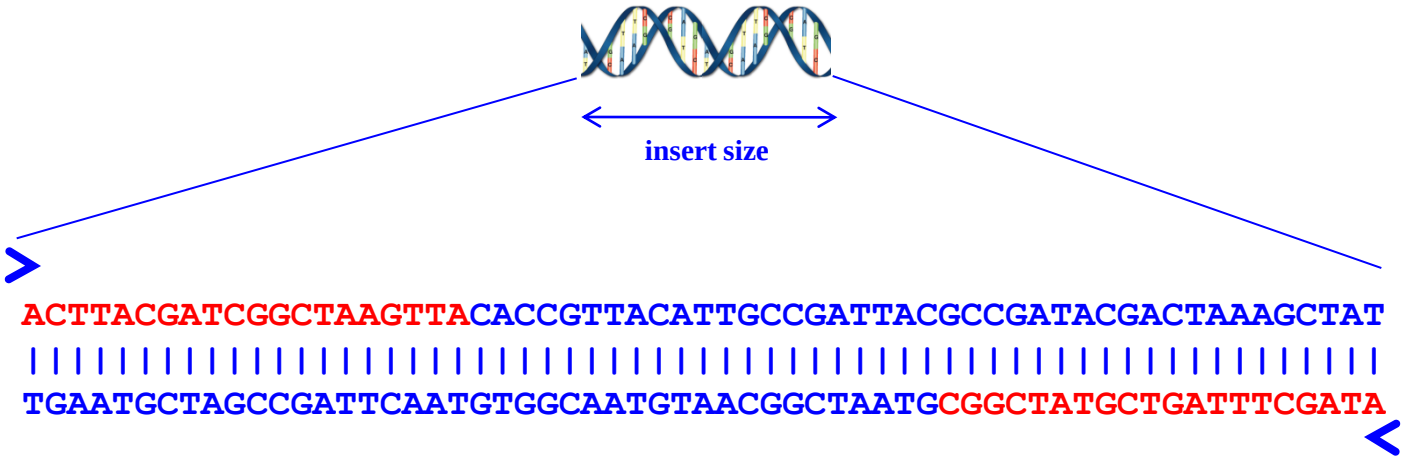
Outline of the Talk:

- *Sequence Reconstruction and Euler Path*
- *Read Clustering based Method*
- *Phusion2 – the Assembly Pipeline*
- *The Devil genome project*
- *Genome sequencing and size estimation*
- *Assigning contigs to individual chromosomes*
- *Devil cancer genome assemblies*
- *Future work*

NGS Sequencing



Genome



Mate-pair

Read 1: ACTTACGATCGGCTAAGTTA

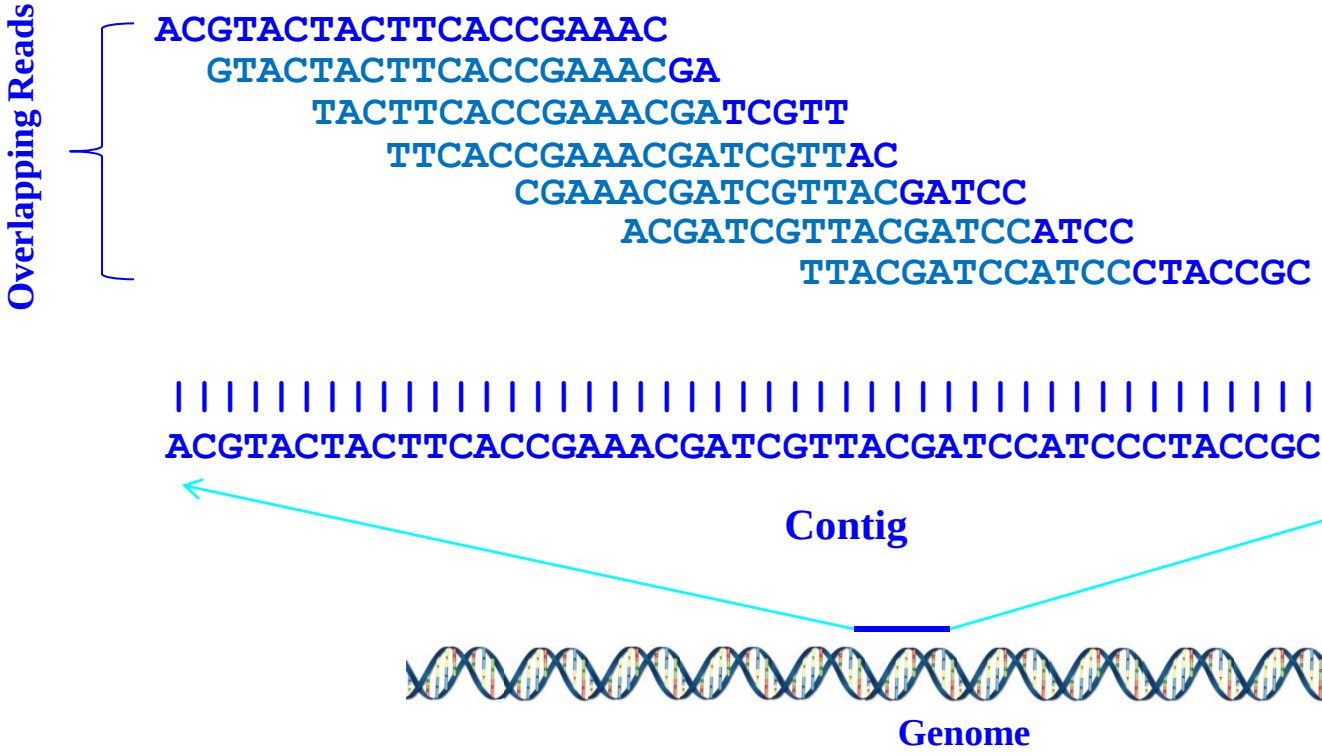
Read 2: ATAGCTTTAGTCGTATCGGC

Read Length: 20

Insert Size: 61

Contigs

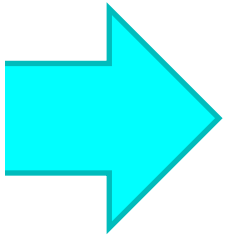
- Contig (from contiguous) is a set of overlapping DNA segments derived from a single genetic source.



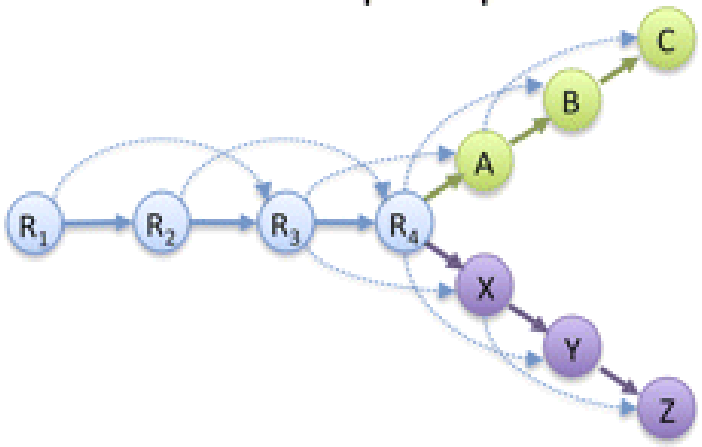
Genome Assembly

Read Layout

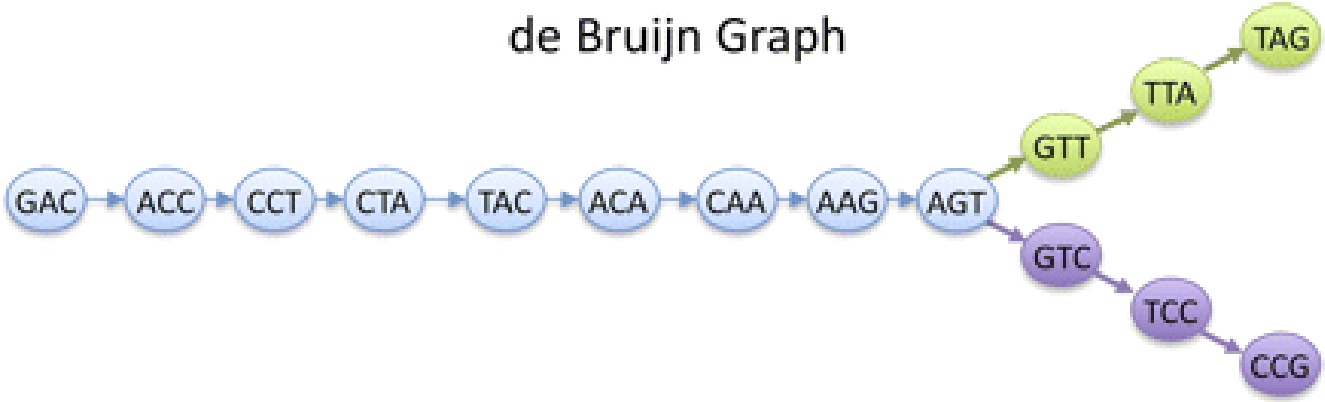
```
R1:   GACCTACA
R2:    ACCTACAA
R3:     CCTACAAG
R4:      CTACAAGT
A:        TACAAGTT
B:         ACAAGTTA
C:          CAAGTTAG
X:        TACAAGTC
Y:         ACAAGTCC
Z:          CAAGTCCG
```



Overlap Graph

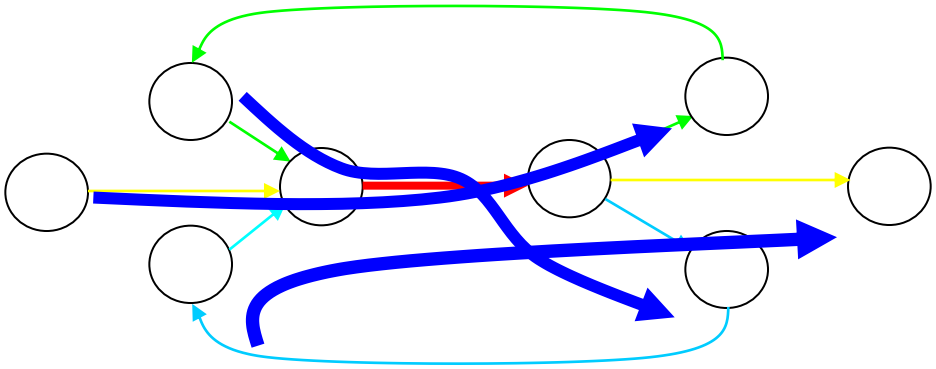


de Bruijn Graph



Sequence Repeat Graph

Sequences

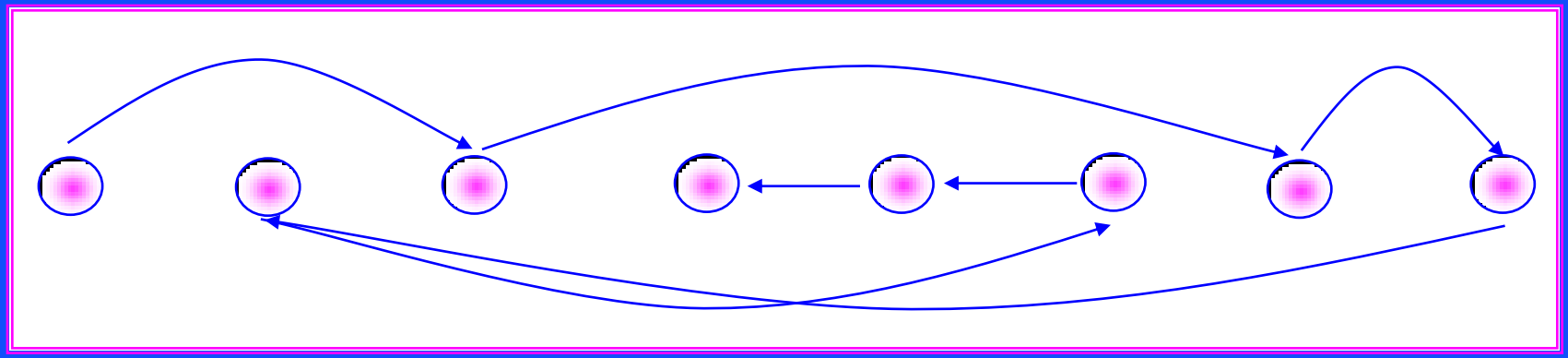


Sequence Reconstruction - Hamiltonian path approach

S=(ATGCAGGTCC)

ATG -> TGC -> GCA -> CAG -> AGG -> GGT -> GTC -> TCC

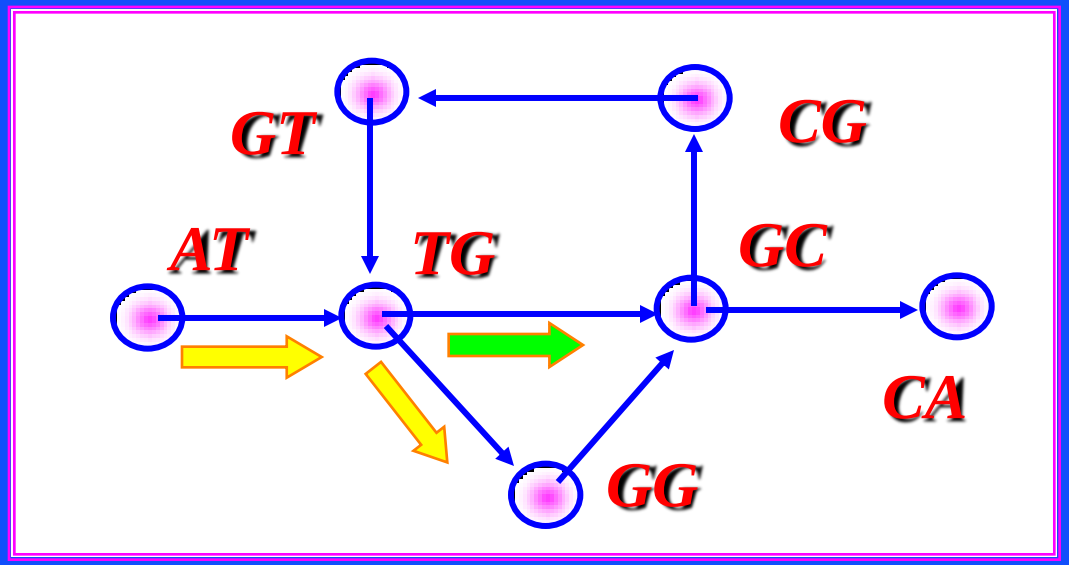
ATG AGG TGC TCC GTC GGT GCA CAG



Vertices: k-tuples from the spectrum shown in red (8);
Edges: overlapping k-tuples (7);
Path: visiting all vertices corresponding to the sequence.

Sequence Reconstruction - Euler path approach

ATG -> TGG -> GGC -> GCG -> CGT -> GTG -> TGC -> GCA



ATGGCGTGCA ATGCGTGGCA

Vertices: correspond to (k-l)-tuples (7);
Edges: correspond to k-tuples from the spectrum (8);
Path: visiting all EDGES corresponding to the sequence.

Kmer Extension & Walk

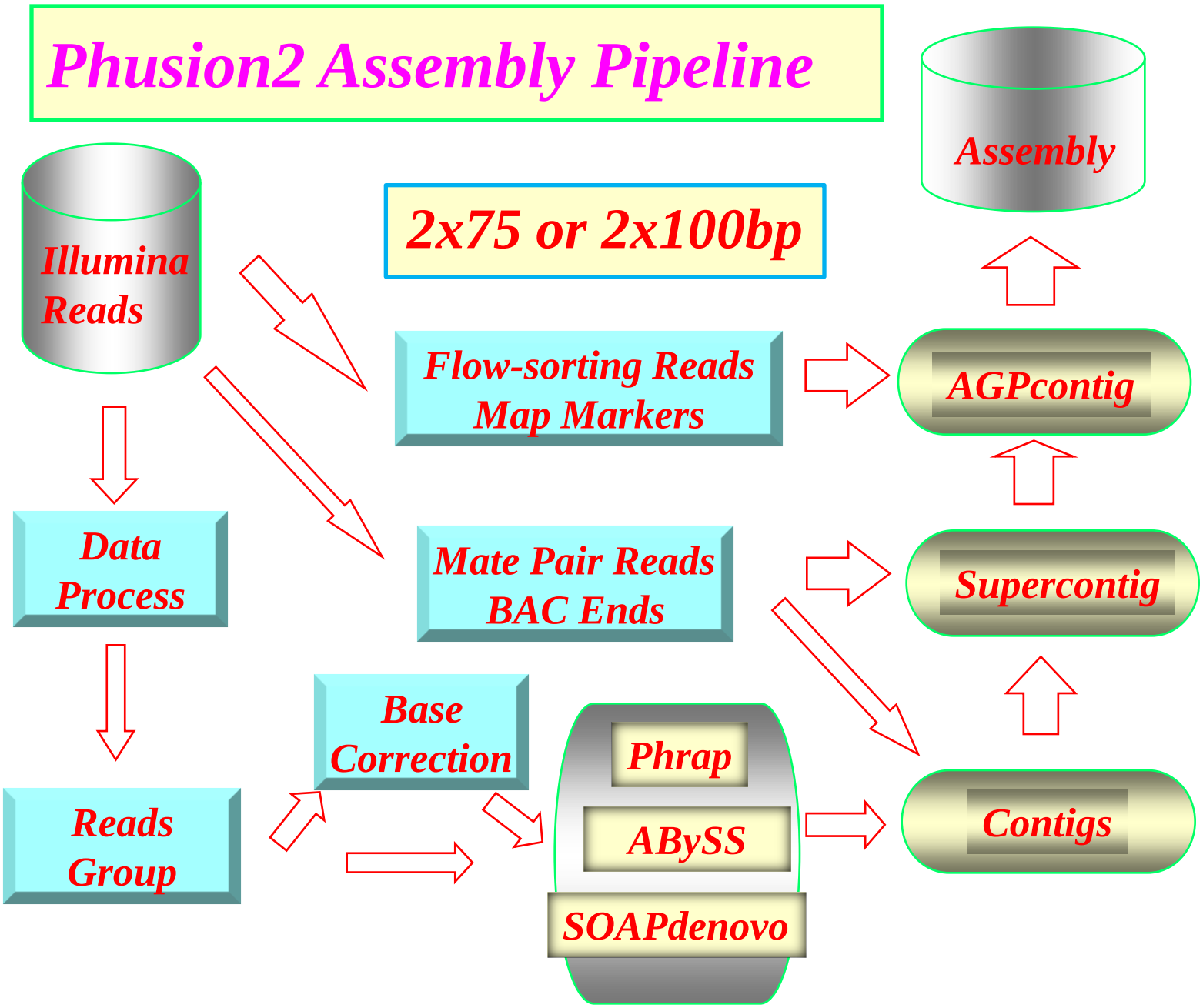
```
contig: 1 108 IL2_33_8_1_571_2
============================
10 1 1  IL2_33_8_46_612_165                  GAAAAGTGGGCTGAGTGGCTGaTgTTCTTgGatgcgggg  w1: 200379443086 801517772347 649109 TTCTT
10 2 3  IL2_33_8_30_108_252                gtGAaAaGtGGGCtGAGtGGCtGATGTTCTTGGATGCGG    w2: 200379443086 801517772347 649109 TTCTT
10 1 4  IL2_33_8_100_266_829              TTTGAAAAGTGGGCTGAGTGGCTGaTGTTcTtggattcg     w1: 200379443086 801517772347 649109 TTNTN
10 1 5  IL2_33_8_102_107_918             ATTTGAAAAGTGGGCTGAGtGGCTGatgtTtttggatct      w1: 200379443086 801517772347 649109 NTNNN
10 2 5  IL2_33_8_40_950_214              AtttGAaaAGTGGGCtGaGtGGCTGATGTTCTTGGATGC      w2: 200379443086 801517772347 649109 TTCTT
10 2 7  IL2_33_8_130_885_970           aaatttgaaaagtGggCtGAGTGGCTGATGTTCTTGGAT         w2: 200379443086 801517772347 649109 TTCTT
10 2 11 IL2_33_8_118_146_767         tAgtAaattggAAAaGTggGCTGAGTGGCTGATGTTCTT         w2: 200379443086 801517772347 649109 TTCTT
10 1 12 IL2_33_8_124_94_829         CTAGTAAATTTGAAAAGTGGGCTGaGTgGcTGatgttct          w1: 200379443086 801517772347 649109 NNNNN
10 2 14 IL2_33_8_7_465_656        gACTAgtaAATTtGAAAAGtGGGCTGAGTGGCTGATGTT           w2: 200379443086 801517772347 649109 TTNNN
10 2 14 IL2_33_8_1_571_251        GaCTAGTAAAtttGAAAAGtGGGCtGAGTGGCTGATGTT           w2: 200379443086 801517772347 649109 TTNNN
-------------------------------------------------------------------------------
1       1 IL2_33_8_1_571_251   NNGACTAGTAAATTTGAAAAGTGGGCTGAGTGGCTGATGTTCTTGGATGCGGGG 10 A
max: 0 649109 1 10 0 0
============================ 2 10
12 1 0  slxa_0011_8_0069_7881                aAAGTGGGCTGAGTGGCTGaTGTTcTTgGaTGgGgg    w1: 801517772347 1325340595487 986799 TNTTN
12 2 1  IL2_33_8_190_796_254               gaaagtgggCtgaGtGGCTGATGTTcTTGGAtGCGGTGG   w2: 801517772347 1325340595487 986799 TNTTG
12 1 2  IL2_33_8_46_612_165                GAAAAGTGGGCTGAGTGGCTGaTgTTCTTgGatgcgggg   w1: 801517772347 1325340595487 986799 TCTTN
12 2 4  IL2_33_8_30_108_252              gtGAaAaGtGGGCtGAGtGGCtGATGTTCTTGGATGCGG     w2: 801517772347 1325340595487 986799 TCTTG
12 1 5  IL2_33_8_100_266_829            TTTGAAAAGTGGGCTGAGTGGCTGaTGTTcTtggattcg     w1: 801517772347 1325340595487 986799 TNTNN
12 1 6  IL2_33_8_102_107_918           ATTTGAAAAGTGGGCTGAGtGGCTGatgtTtttggatct      w1: 801517772347 1325340595487 986799 TNNNN
12 2 6  IL2_33_8_40_950_214            AtttGAaaAGTGGGCtGaGtGGCTGATGTTCTTGGATGC      w2: 801517772347 1325340595487 986799 TCTTG
12 2 8  IL2_33_8_130_885_970         aaatttgaaaagtGggCtGAGTGGCTGATGTTCTTGGAT         w2: 801517772347 1325340595487 986799 TCTTG
12 2 12 IL2_33_8_118_146_767       tAgtAaattggAAAaGTggGCTGAGTGGCTGATGTTCTT         w2: 801517772347 1325340595487 986799 TCTTN
12 1 13 IL2_33_8_124_94_829       CTAGTAAATTTGAAAAGTGGGCTGaGTgGcTGatgttct          w1: 801517772347 1325340595487 986799 NNNNN
12 2 15 IL2_33_8_1_571_251      GaCTAGTAAAtttGAAAAGtGGGCtGAGTGGCTGATGTT           w2: 801517772347 1325340595487 986799 TNNNN
12 2 15 IL2_33_8_7_465_656      gACTAgtaAATTtGAAAAGtGGGCTGAGTGGCTGATGTT           w2: 801517772347 1325340595487 986799 TNNNN
-------------------------------------------------------------------------------
2       1 IL2_33_8_1_571_251   NGACTAGTAAATTTGAAAAGTGGGCTGAGTGGCTGATGTTCTTGGATGCGGGGGN 12 A
max: 0 986799 1 12 0 0
============================ 3 12
13 1 0  IL2_33_8_7_465_656                   AACATCAGCCACTCAGCCCaCTTTTCaAATTtacTAGTc
13 1 0  IL2_33_8_1_571_251                   AACATCAGCCACTCaGCCCaCTTTTCaaaTTTACTAGtC
13 2 2  IL2_33_8_124_94_829                agaacatCAgCcACtCAGCCCACTTTTCAAATTTACTAG   w3: 1325340595487 12824284357565 4962647 NNNNN
13 1 3  IL2_33_8_118_146_767              AAGAACATCAGCCACTCAGCcCACtTTTccaattTacTa    w4: 1325340595487 12824284357565 4962647 GAANN
13 1 7  IL2_33_8_130_885_970            ATCCAAGAACATCAGCCACTCaGccCacttttcaaattt      w4: 1325340595487 12824284357565 4962647 GAACC
13 1 8  slxa_0011_8_0043_10001          cATCCAAGAACATCAGCCACTCAGCCCaCtTCTCAc         w4: 1325340595487 12824284357565 4962647 GAACC
13 2 9  IL2_33_8_102_107_918           agatccaaaAacatCAGCCaCTCAGCCCACTTTTCAAAT      w3: 1325340595487 331335148871 375421 NNNNN
13 1 9  IL2_33_8_40_950_214            GCATCCAAGAACATCAGCCaCtCaGCCCACTttTCaaaT      w4: 1325340595487 12824284357565 4962647 GAACC
13 2 10 IL2_33_8_100_266_829          cgaatccaAgAACAtCAGCCACTCAGCCCACTTTTCAAA      w3: 1325340595487 12824284357565 4962647 NANNN
13 1 11 IL2_33_8_30_108_252          CCGCATCCAAGAACATCaGCCaCTCaGCCCaCtTtTCac       w4: 1325340595487 12824284357565 4962647 GAACC
13 2 12 slxa_0011_8_0069_7881       ccCcCAtCcAAgAACAtCAGCCACTCAGCCCACTTt          w3: 1325340595487 12824284357565 4962647 NAANC
13 2 13 IL2_33_8_46_612_165        ccccgcatCcAAGAAcAtCAGCCACTCAGCCCACTTTTC       w3: 1325340595487 12824284357565 4962647 GAANC
13 1 14 IL2_33_8_190_796_254      CCACCGCaTCCAAgAACATCAGCCaCtcaGcccactttc        w4: 1325340595487 12824284357565 4962647 NAACC
-------------------------------------------------------------------------------
3       1 IL2_33_8_1_571_251   NNCCCCCGCATCCAAGAACATCAGCCACTCAGCCCACTTTTCAAATTTACTAGTC 11 A
max: 0 375421 1 1 0 0
max: 1 4962647 1 10 1 1
============================ 4 13
13 2 0  IL2_33_8_15_630_886                    AGTGgGCtgAGTGGCTGATGTTCTTGGATGCGGTGGATG w2: 12824284357565 8878926809425 3711226 TTGGA
13 1 0  IL2_33_8_182_360_770                   AGtGggCTGAGtggctgatgttcttgggatagggggagc w1: 12824284357565 8878926809425 3711226 NNNNN
13 1 2  slxa_0011_8_0069_7881                aAAGTGGGCTGAGTGGCTGaTGTTcTTgGaTGgGgg     w1: 12824284357565 8878926809425 3711226 TTNGN
13 2 3  IL2_33_8_190_796_254                gaaagtgggCtgaGtGGCTGATGTTcTTGGAtGCGGTGG   w2: 12824284357565 8878926809425 3711226 TTGGA
```

Base Quality to Filter Base Errors

```
29 2 13 IL2_33_8_116_24                                                                                    964 10129248 NANNN
29 1 13 IL2_33_8_92_670                                                                                    964 10129248 TACCA
--------------------------------------------------------------------------------
63      2 IL2_33_8_1_920_483   NNNAAAGCAAAACCATAGCCGATAAAGAGGTTGAGGTCCAAGAGTTCATTGGATA 26 G
max: 0 10129248 2 26 0 0
============================== 64 29
27 1 0  IL2_33_8_83_322_97                    GACCTCAACCTCTTTATCGGCTattgATTtcTcCAtaTC w1: 36782494135964 6392488188531 3044891 NNNAT
27 1 2  slxa_0011_8_0008_13230              tGGACCTCaACCTCTTTATCGGCTatgGtTTTgttt      w1: 36782494135964 6392488188531 3044891 NNGNT
27 1 2  slxa_0011_8_0016_3358               tGGACCTCAACCTCTTTATCGGCTATGGTTTTgTTT      w1: 36782494135964 6392488188531 3044891 TGGTT
27 2 4  IL2_33_8_92_670_287              catGgaccTCAACCtcTTTAtCGGCTATGGTTTTGCTAT     w2: 36782494135964 6392488188531 3044891 TGGTT
27 2 4  slxa_0011_8_0015_3638             CacGGACCTCAACCTCTTTATCGGCTATGGTTTTGc       w2: 36782494135964 6392488188531 3044891 TGGTT
27 1 4  IL2_33_8_116_24_685               CTTGGACCTCAACCTCTTTATCGgcTaTggttttttctcT   w1: 36782494135964 6392488188531 3044891 TNNNN
27 2 5  slxa_0011_8_0048_11543           tCTTGgACCTCAACCTCTTTATCGGCTATGGTTTTg       w2: 36782494135964 6392488188531 3044891 TGGTT
27 1 5  slxa_0011_8_0017_12919           tCTTGGACCTCAACCTCTTTATCGGCTAttgtTTtt       w1: 36782494135964 6392488188531 3044891 NNNNT
27 1 7  IL2_33_8_99_283_956            ACTCTTGGACCTCAACCTCTTTATCGGCTAtaGttTTtt     w1: 36782494135964 6392488188531 3044891 NNGNN
27 2 7  IL2_33_8_127_78_352            actctgggACctCAaCCtCtTTATCGGCTATGGTTTTGC     w2: 36782494135964 6392488188531 3044891 TGGTT
27 2 8  IL2_33_8_180_283_791           AACtCttggACCtCAACCTCTTTATCGGCTATGGTTTTG      w2: 36782494135964 6392488188531 3044891 TGGTT
27 1 8  IL2_33_8_190_507_900           AACTCTTGGACCTCAAcCTCTTTATcGGCTAatgttttT     w1: 36782494135964 6392488188528 3044890 NNNNN
27 1 9  slxa_0011_8_0068_4180         gAACTCTTGGACCTCAACCTCTTTATCGGCTATGGt         w1: 36782494135964 6392488188531 3044891 TGGNN
27 2 9  IL2_33_8_86_949_588           gagCtCttGGaCCTcAACCTCTTTATCGGCTATGGTTTT      w2: 36782494135964 6392488188531 3044891 TGGTT
27 1 9  IL2_33_8_156_75_481           GAACTCTTGGACCTCAACCTCTTTATCGgCTAttgtttT      w1: 36782494135964 6392488188531 3044891 NNNNN
27 1 9  IL2_33_8_195_626_572          GAACTCTTGGACCTCAACCTCTTTATCGGCTaatgtTTT      w1: 36782494135964 6392488188528 3044890 NNNNT
27 1 11 slxa_0011_8_0018_3598        aTGAACTCTTGGACCTCAACCTCTTTATCGGCTatg         w1: 36782494135964 6392488188531 3044891 NNNNN
27 2 11 slxa_0011_8_0034_10003       AtgAaCtCttGGACCTCAACCTCTTTATCGGCTATg         w2: 36782494135964 6392488188531 3044891 TNNNN
27 1 11 IL2_33_8_128_890_473         ATGAACTCTTGGACCTCAACCTCTTTATCGGCtAtgtaT      w1: 36782494135964 6392488188531 3044891 NNNNT
27 2 11 IL2_33_8_16_310_346          AtGAaCtCttGGACCtCAACCTcTTTATCGGCTATGGTT      w2: 36782494135964 6392488188531 3044891 TGGTT
27 1 12 IL2_33_8_187_928_742        AATGAACTCTTGGACCTCAAcCTCTTTATcggcTatttt       w1: 36782494135964 6392488188531 3044891 NNNNN
27 1 12 IL2_33_8_31_361_610         AATGAACTCTTGGACCTCaaCCTCTTTATCGgCTaatct       w1: 36782494135964 6392488188528 3044890 NNNNN
27 1 12 IL2_33_8_39_824_524         AATGAACTCTTGGACCTCAACCTCtTTATCgGCTaagtt       w1: 36782494135964 6392488188528 3044890 NNNNN
27 2 13 IL2_33_8_182_476_783       CtatGaactCtTggACCtCAACCTCTTTATCGGCTATGG        w2: 36782494135964 6392488188531 3044891 TGGNN
27 1 13 IL2_33_8_32_502_869        CAATGAACTCTTGGACCTCAACCTCTTTATCGGCTaatg        w1: 36782494135964 6392488188528 3044890 NNNNN
27 2 15 IL2_33_8_187_932_219     acCAatgaaCtCttGgaCCTCAACCTCTTTATCGGCTAT         w2: 36782494135964 6392488188531 3044891 TNNNN
27 2 16 IL2_33_8_165_814_103    AtcCAAtGAACTCTTGGACCTCAACCTCTTTATCGGCTA
--------------------------------------------------------------------------------
64      2 IL2_33_8_1_920_483   AACCAATGAACTCTTGGACCTCAACCTCTTTATCGGCTATGGTTTTGCTCTTATC 26 G
max: 0 3044890 2 5 0 0
max: 5 3044891 2 21 5 5
repeat: 26 0.807692 0.238095 5 64
Go: 13 0
NKM: 5 26 2.000000 0.000000 NNTTTTTNNTTNTTNNNTNTNNNTNT 13 0
rate-c: 5 13 0 2.000000 26
pairs: 2 0 0
rate: 21 5 4.200000 0 3044891
============================== 65 27
21 1 1  IL2_33_8_83_322_97                     GACCTCAACCTCTTTATCGGCTattgATTtcTcCAtaTC  w1: 6392488188531 3459798052746 1903672 NNATT
21 1 3  slxa_0011_8_0016_3358                tGGACCTCAACCTCTTTATCGGCTATGGTTTTgTTT       w1: 6392488188531 21051984097162 7403021 GGTTT
21 1 3  slxa_0011_8_0008_13230               tGGACCTCaACCTCTTTATCGGCTatgGtTTTgttt       w1: 6392488188531 21051984097162 7403021 NGNTT
21 2 5  IL2_33_8_92_670_287                catGgaccTCAACCtcTTTAtCGGCTATGGTTTTGCTAT      w2: 6392488188531 21051984097162 7403021 GGTTT
21 1 5  IL2_33_8_116_24_685                CTTGGACCTCAACCTCTTTATCGgcTaTggttttttctcT      w1: 6392488188531 21051984097162 7403021 NNNNN
21 2 5  slxa_0011_8_0015_3638              CacGGACCTCAACCTCTTTATCGGCTATGGTTTTGc        w2: 6392488188531 21051984097162 7403021 GGTTT
:
```

Read Pairs in Repeat Junctions

```
max: 23 3000904 5 23 1 0
============================== 1806 24
26 1 0  IL2_33_8_82_544_274                  ACCGCTTCTCAAAGTTAGTGTCACCATCTCAccGCAGtG w1: 6182404869044 24729619476177 8160638 CCATC
26 1 0  IL2_33_8_25_905_631                  ACCGCTTCTCAAAGTTAGTGTCACCATCtCAGCGCAgtG w1: 6182404869044 24729619476177 8160638 CCATC
26 1 0  IL2_33_8_162_151_903                 ACCGCTTCTCAAAGTTAGTGTCAACaTCTCAgCGCagag w1: 6182404869044 24729619476176 8160637 ACNTC
26 1 1  slxa_0011_8_0001_11718               cACCGCTTCTCAAAGTTAGTGTCAACaTCTCaGCGC     w1: 6182404869044 24729619476176 8160637 ACNTC
26 1 2  IL2_33_8_189_106_58                 GCACCGCTTCTCAAAGtTAGTGTCaCCaTCTCagcgccg  w1: 6182404869044 24729619476177 8160638 CCNTC
26 1 2  IL2_33_8_99_681_567                 GCACCGCTTCTCAAAGtTAgTGTCAACATCTCaGCGCaG  w1: 6182404869044 24729619476176 8160637 ACATC
26 1 4  IL2_33_8_146_466_847               CAGCACCGCTTCTCAAAGTTAGTGTCACCATCtcagcgc    w1: 6182404869044 24729619476177 8160638 CCATC
26 2 4  IL2_33_8_67_102_681                 cAccaccGctTCtcAAAGTTAGTGTCACCATCTCAGCGC    w2: 6182404869044 24729619476177 8160638 CCATC
26 1 4  IL2_33_8_109_251_575                CAGCACCGCTTCTCAAAGTTAGTGTCaaCATCTCAgcgC    w1: 6182404869044 24729619476176 8160637 NCATC
26 2 5  slxa_0011_8_0061_6558              CcagcacCGCTTCTCAAAGTTAGTGTCACCATCTCa        w2: 6182404869044 24729619476177 8160638 CCATC
26 1 7  IL2_33_8_58_33_385                GCCCAGCACCGCTTCTCaaaGtTAGtGtCaacatctcag     w1: 6182404869044 24729619476176 8160637 NNNNN
26 2 7  IL2_33_8_142_159_871              gcccagcaCCgCTTCTCAAAGTTAGTGTCACCATCTCAG     w2: 6182404869044 24729619476177 8160638 CCATC
26 1 7  slxa_0011_8_0029_3657             gCCCAGCACCGCtTCTCAAAGtTAgtGTCACcATct        w1: 6182404869044 24729619476177 8160638 CNATN
26 1 7  IL2_33_8_86_279_570               GCCCAGCACCGCTTCTCaaaGtTAGTGTCAACaTCTCcg     w1: 6182404869044 24729619476176 8160637 ACNTC
26 1 8  slxa_0011_8_0060_1025            aGCCCAGCACCGCTTCTCAAAGTTAGTGTCACCatc        w1: 6182404869044 24729619476177 8160638 CCNNN
26 1 8  slxa_0011_8_0016_9354            aGCCCAGCACCGCTTCTCAAagTTAGtGTCaCcatC        w1: 6182404869044 24729619476177 8160638 CNNNC
26 1 9  IL2_33_8_154_131_74             AAGCCCAGCACCGCTTCTCAaAGTTaGTGTCAccatCTc     w1: 6182404869044 24729619476177 8160638 NNNNC
26 2 10 IL2_33_8_161_605_406            AAagCccagCaCCGCTTCTCAAAGTTAGTGTCAACATCT     w2: 6182404869044 24729619476176 8160637 ACATC
26 2 11 IL2_33_8_3_768_128             taacgcacagcacCGCtTCTCAAAGTTAGTGTCACCATC     w2: 6182404869044 24729619476177 8160638 CCATC
26 1 11 IL2_33_8_56_938_133            TCAAGCCCaGCACCGCTTCTCAAAGtTAGTGtCaCCAtc     w1: 6182404869044 24729619476177 8160638 CCANN
26 1 12 IL2_33_8_187_563_450           CTCAAGCcCaGCACCGCTTCTCaaAGttaGtgTcaccat     w1: 6182404869044 24729619476177 8160638 NNNNN
26 1 13 IL2_33_8_43_994_259            GCTCAAGCCCAGCACCGCTTCTCaaAGtTAGTGTCaCCc     w1: 6182404869044 24729619476177 8160638 CCNNN
26 1 14 IL2_33_8_156_822_939          GGCTCAAGCCCAGCACCGCTTCtCAaAGtTaGtgtcacc      w1: 6182404869044 24729619476177 8160638 NNNNN
26 2 14 IL2_33_8_168_237_83           gaCtcAcgcccAGcacCGCTtCTCAAAGTTAGTGTCACC      w2: 6182404869044 24729619476177 8160638 CCNNN
26 1 16 IL2_33_8_176_937_717    tGGGCTCaaGCcCaGCACcgCttctCaaagttAgtGtca
26 2 16 IL2_33_8_153_548_484    ggagcacAagcccagCACCGCTTCTCAAAGTTAGTGTCA
--------------------------------------------------------------------------------
1806    5 IL2_33_8_1_712_298    GGGGCTCAAGCCCAGCACCGCTTCTCAAAGTTAGTGTCACCATCTCAGCGCAGTG 24 A
max: 0 8160637 5 7 0 0
max: 7 8160638 5 17 7 7
repeat: 24 0.708333 0.411765 7 1806
NKM: 0 24 0.583333 0.208333 CCAACACCNCNCCACCNACCNCNC 14 5
NKM: 1 24 2.000000 0.000000 CCCCCCCCCCNCNCCNNCCCNCNC 18 0
rate-c: 1 18 0 2.000000 24
NKM: 2 24 2.000000 0.000000 AANNNAAAAANAANNNNAAANNNN 12 0
rate-c: 2 12 0 2.000000 24
NKM: 3 24 2.000000 0.000000 TTTTTTTTTTNTTTNNNTTNNNNN 15 0
rate-c: 3 15 0 2.000000 24
NKM: 4 24 2.000000 0.000000 CCCCCCCCCCNCNCNCCCCNNNNN 16 0
rate-c: 4 16 0 2.000000 24
break2: 0
pass: 0  IL2_33_8_86_279_570            AAagCccagCaCCGCTTCTCAAAGTTAGTGTCAACATCT 6182404869044 8160637 14 18
pass2: 0 1329 0 1806
pass: 1  IL2_33_8_161_605_406              CAGCACCGCTTCTCAAAGTTAGTGTCaaCATCTCAgcgC 6182404869044 8160637 14 18
:
```


Phusion2 Assembly Pipeline
Assembly
Illumina Reads
2x75 or 2x100bp
Flow-sorting Reads
Map Markers
AGPcontig
Data Process
Mate Pair Reads
BAC Ends
Supercontig
Base Correction
Phrap
ABySS
SOAPdenovo
Reads Group
Contigs

Kmer Word Hashing

ATGGCGTGCAGTCCATGTTCGGATCA

ATGGCGTGCAGT

TGGCGTGCAGTC

GGCGTGCAGTCC

GCGTGCAGTCCA

CGTGCAGTCCAT

Contiguous Base Hash K = 12

ATGGCGTGCAGTCCATGTTCGGATCA

ATGGGCAGATGT

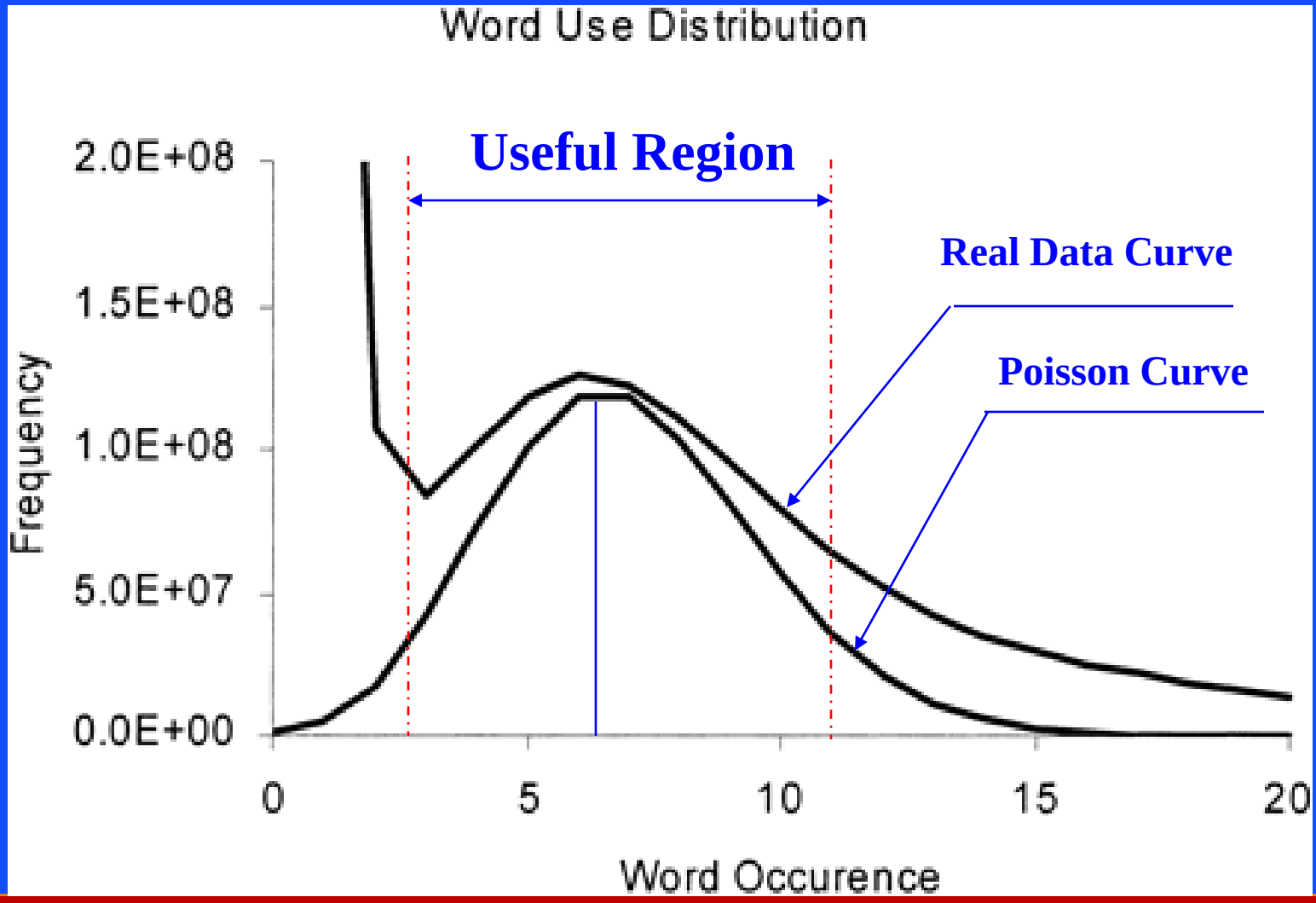
TGGCCAGTTGTT

GGCGAGTCGTTC

GCGTGTCCTTCG

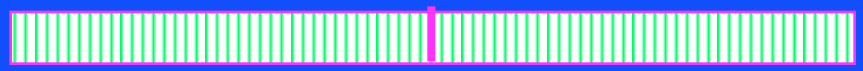
Gap-Hash 4x3

Word use distribution for the mouse sequence data at ~7.5 fold



Sorted List of Each *k*-Mer and Its Read Indices

High bits	Low bits
ACAGAAAAGC	10h06.p1c
ACAGAAAAGC	12a04.q1c
ACAGAAAAGC	13d01.p1c
ACAGAAAAGC	16d01.p1c
ACAGAAAAGC	26g04.p1c
ACAGAAAAGC	33h02.q1c
ACAGAAAAGC	37g12.p1c
ACAGAAAAGC	40d06.p1c
ACAGAAAAGG	16a02.p1c
ACAGAAAAGG	20a10.p1c
ACAGAAAAGG	22a03.p1c
ACAGAAAAGG	26e12.q1c
ACAGAAAAGG	30e12.q1c
ACAGAAAAGG	47a01.p1c



64 -2k *2k*

Relation Matrix: R(i,j) – number of kmer words shared between read i and read j

	1	2	3	4	5	6	… j …	N
1		41	0	0	0	0		
2	41		37	0	0	0		
3	0	37		0	22	0		
4	0	0	0		0	27		
5	0	0	22	0		0		
6	0	0	0	27	0			
i								
N								

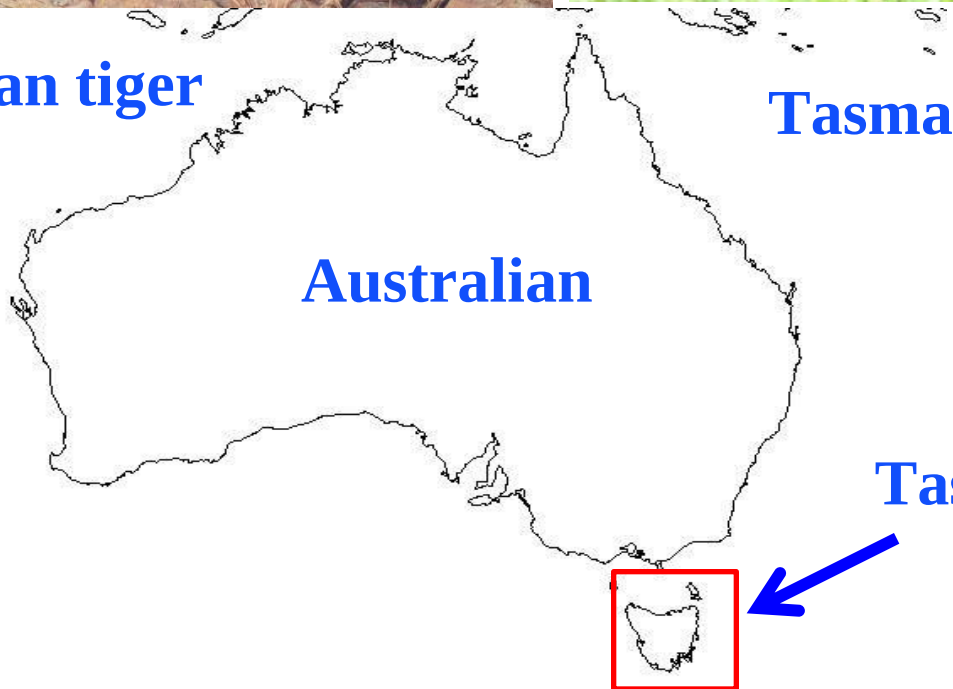
Group 2: (4,6)

Group 1: (1,2,3,5)

R(i,j)

Tasmanian tiger

Tasmanian devil



SAVE THE TASMANIAN DEVIL PROGRAM

SAVE THE TASMANIAN DEVIL.

Devil Facial Tumour Disease threatens the existence of this internationally-recognised icon.
In some areas more than 90% of the Tasmanian devil population has been wiped out. Read more...

DONATE TO THE APPEAL

WELCOME

This site is your primary source for authoritative, up to date information on Devil Facial Tumour Disease (DFTD). We will keep you informed of what is being done to save the Tasmanian devil and how you can help.

Read about the Roadkill Project

Latest News (also see our newsletters)

The devil inside

Where would you least expect to find a Tasmanian devil...now, where in your KITCHEN would you least expect to find a Tasmanian devil?

Published: 01/03/2011

Yianni's Tassie devil science project raises awareness in the U.S.

WHO WE ARE

The Save the Tasmanian Devil Program is the official response to the threat to the survival of the Tasmanian devil.

The Program is an initiative of the Australian and Tasmanian governments.

Australian Government

Tasmania

Current Events

Threatened Species Day 2011

Wherever you are!
07/09/2011

Follow us...

facebook twitter

Devil facts

The Insurance Population will

Completed genome is first step to tackling DFTD

Published: 29/09/2010

A draft genome sequence for the Tasmanian devil will be used to find genetic mutations in the Devil Facial Tumour Disease (DFTD), researchers from the Wellcome Trust Sanger Institute and Illumina announced in September 2010.

The results open the door for new research to pick out those specific mutations that drive the cancer and will lay the foundation for ongoing work to trace the spread of disease, said Dr Elizabeth Murchison, a researcher from the Sanger Institute.

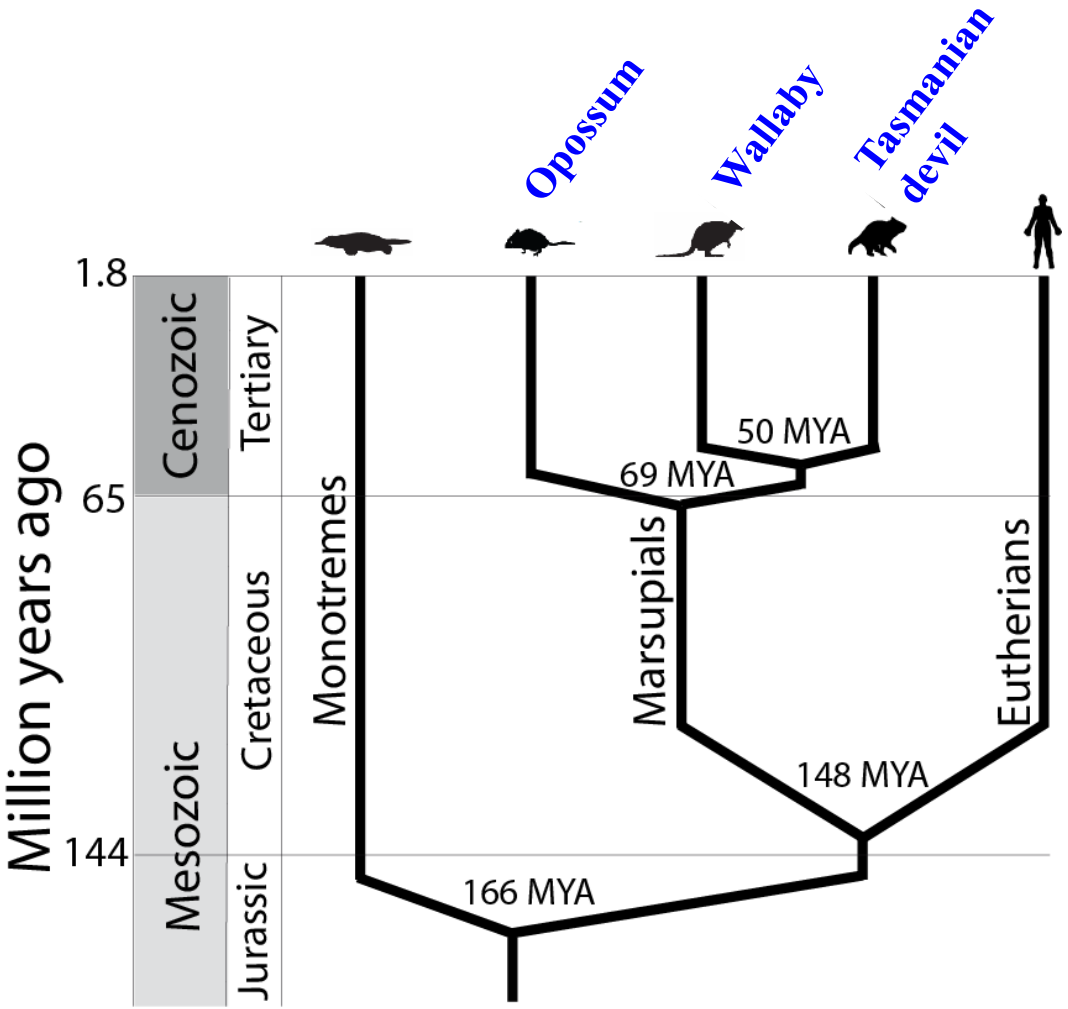
"This sequence is invaluable and comes at a crucial time," she said.

"By comparing our draft sequence with samples taken from many hundreds of devils suffering from this cancer, we can begin to look at the spread of the disease, quite literally, by identifying geographical routes and barriers in its transmission. This knowledge could ultimately shape the ongoing conservation efforts in Tasmania.

"It took 10 years to sequence the draft human genome; the devil took just two months using this new technology. We are entering a new era where genome sequencing can be applied to some of our most pressing problems in real time."

As well as producing the draft genome sequence for a healthy Tasmanian devil, the team has

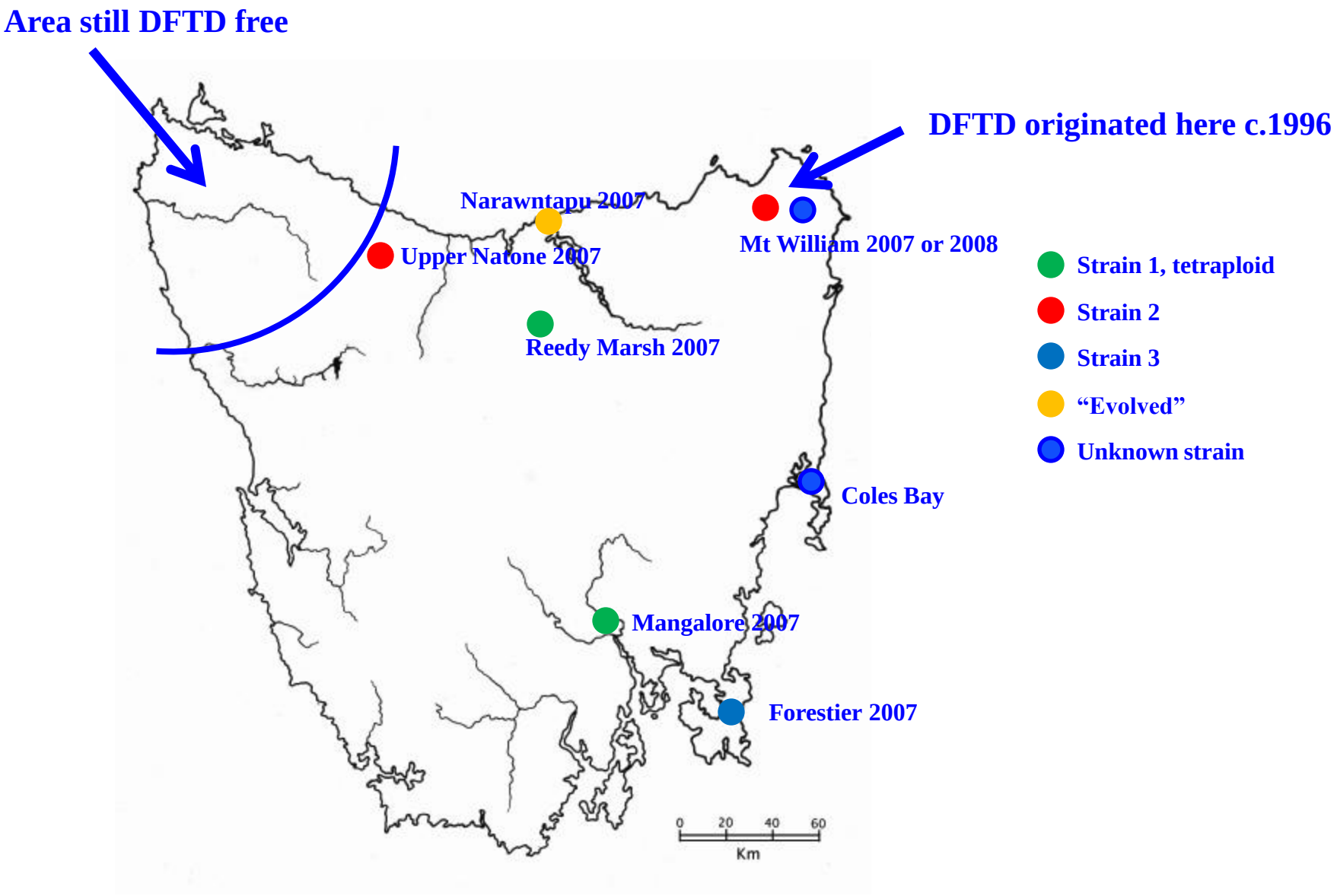
Tasmanian devil



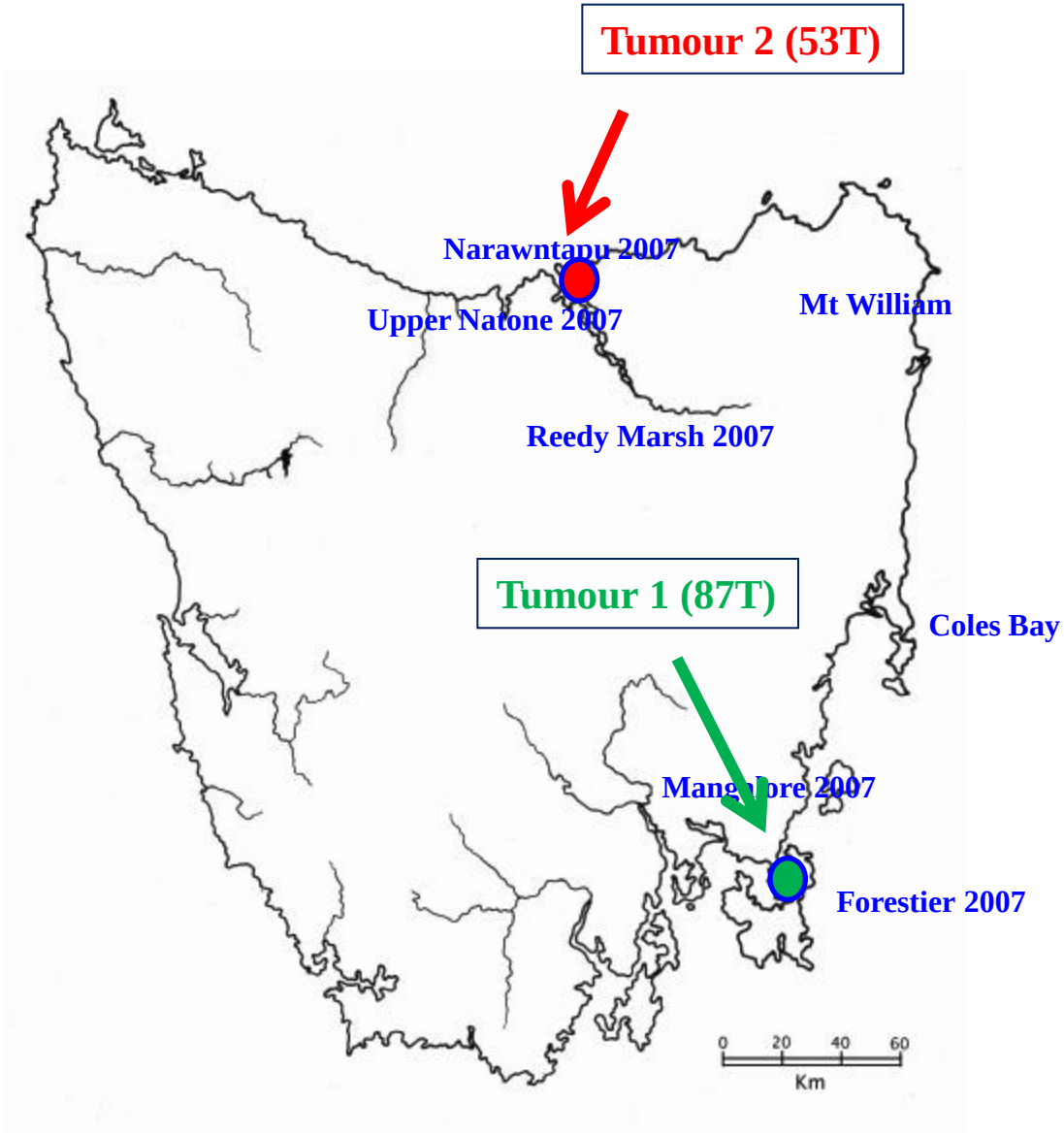
Tasmanian devil facial tumour disease (DFTD)

- Transmissible cancer characterised by the growth of large tumours on the face, neck and mouth of Tasmanian devils
- Transmitted by biting
- Commonly metastasises
- First observed in 1996
- Primarily affects adults >1yr
- Death in 4 – 6 months

DFTD samples for sequencing



Devil Genomes Sequenced

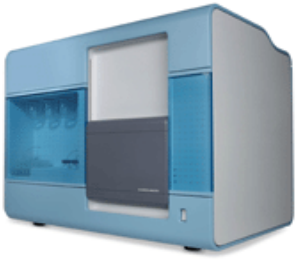


Salem - A female Tasmanian Devil lived Taronga Zoo in Sydney.

Sequencing T. Devil on Illumina: Strategy

Tumour or normal genomic DNA

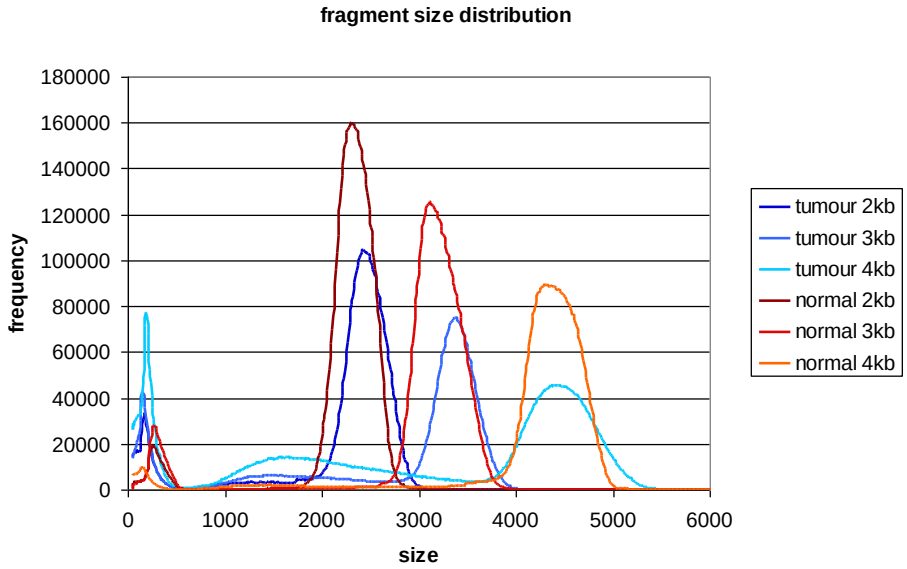
Fragments of defined size
0.5, 2, 5, 7, 8, 10 kb



Sequencing

2x100bp reads short insert

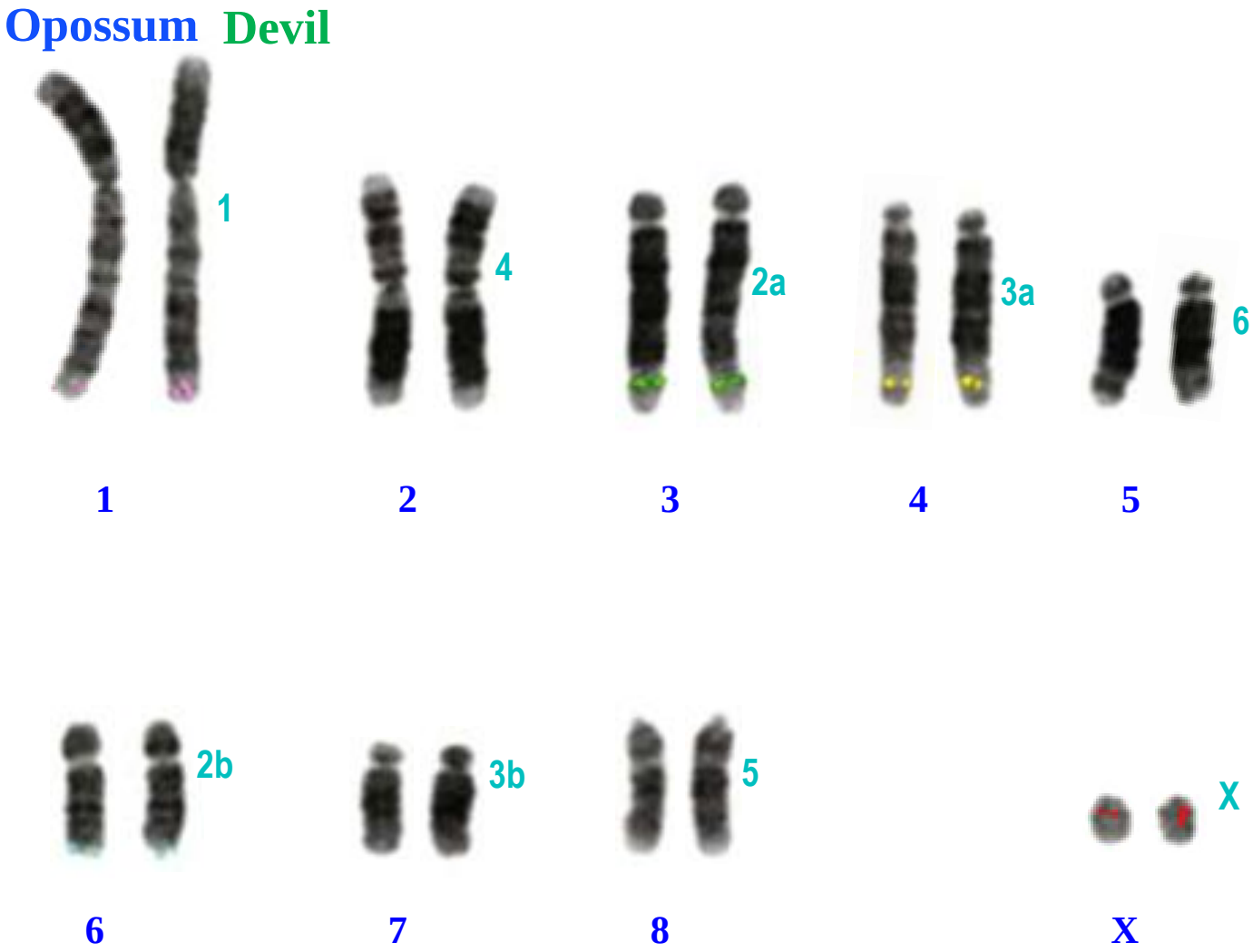
2x50bp mate pairs



Sequencing performed at Illumina

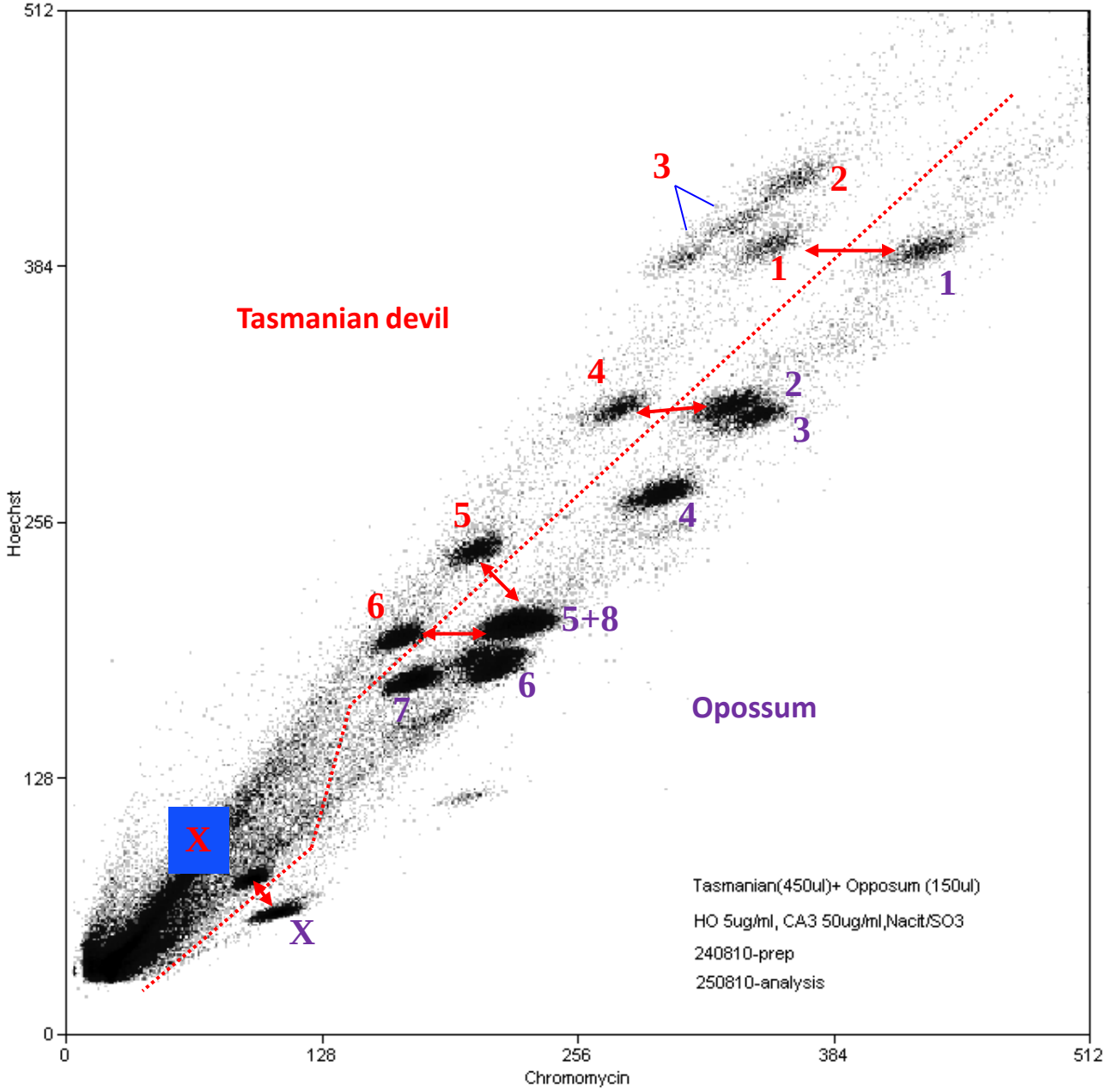
	Salem (91H)	Joey (31H)	Cancer 1 (87T)	Cancer 2 (53T)
Read Coverage	85x	40x	56x	84x

Devil – Opossum Homology Map Based on Hybridisation Results of Devil Paints onto Opossum Chromosomes



Opossum chromosome images were taken from Duke et a. 2007, Chromosome Res 15:361-370

Flow cytometry analysis of chromosomal mixture of devil and opossum



Genome size

	Opossum		Devil
Chr	Seq	FC	FC
1	748	611	571
2	541	484	610
3	526	483	556
4	430	423	450
5	309	321	341
6	245	296	277
7	263	264	
8	308	321	
X	61	116	121
Total	3431	3319	2926

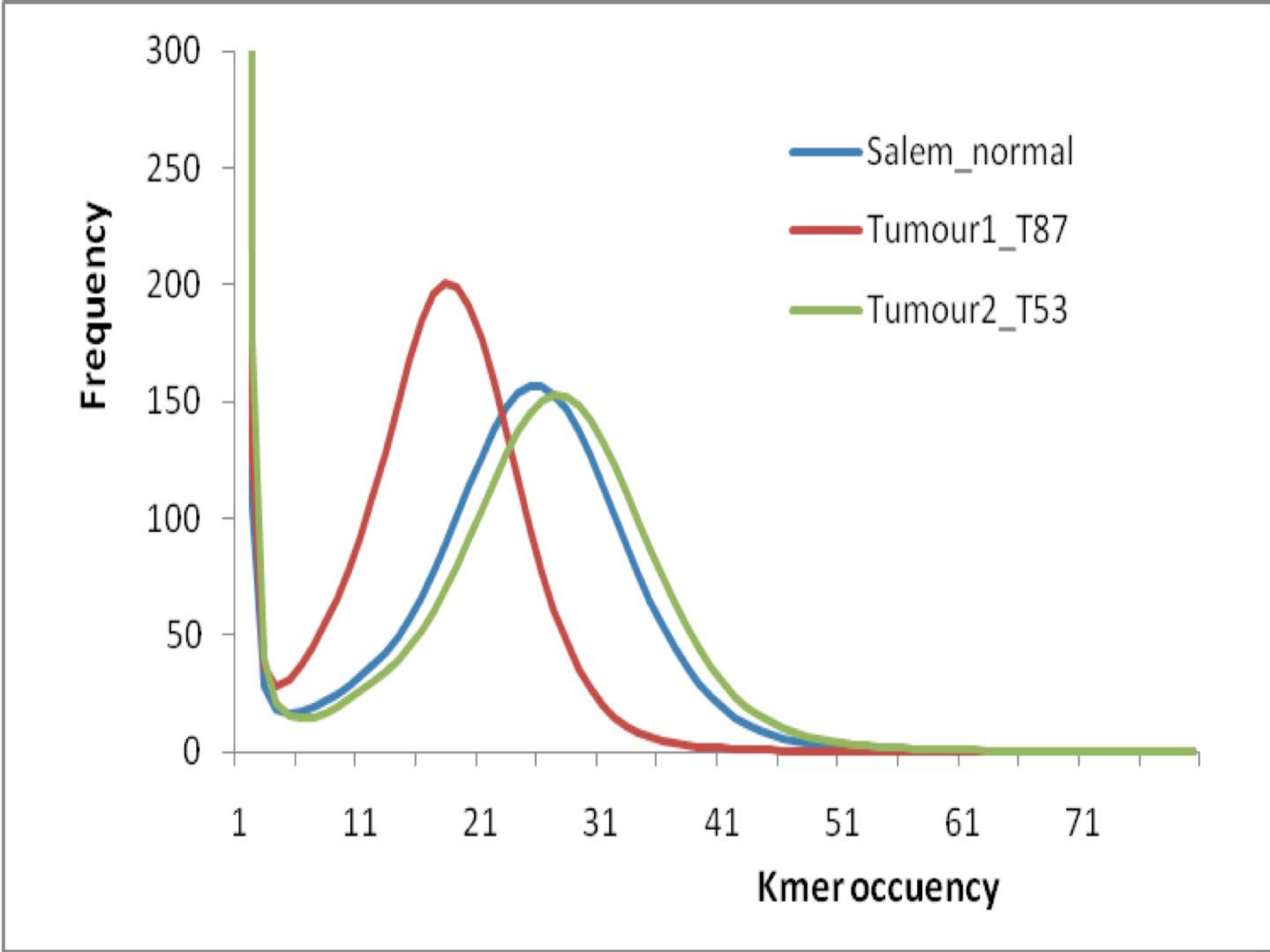

300
250
200
150
100
50
0
Frequency
1
11
21
31
41
51
61
71
Kmer occuency
Salem_normal
Tumour1_T87
Tumour2_T53

Table 1 Run ID, Template names, Number of reads and Chromosome size

4972_1	chr1	IL20_4972:1	19.8	571
4967_1	chr2	IL21_4967:1	20.0	610
4971_1	chr3	IL30_4971:1	21.7	556
4964_1	chr4	IL14_4964:1	7.26	450
4969_1	chr5	IL17_4969:1	7.06	341
4969_2	chr6	IL17_4969:2	8.59	277
4969_3	chrx	IL17_4969:3	9.43	122

Read mapping coefficient:

$$e = Size\_of\_Chr/Num\_reads\_in\_lane$$

Terminal

File Edit View Terminal Tabs Help

```
cigar:A:52 IL30_4971:1:65:14139:10506/2 74 1 - Contig_000301929 10482 10555 + 68 M 74
cigar:A:53 IL30_4971:1:84:11619:8261/1 25 76 + Contig_000301929 10540 10591 + 52 M 52
cigar:A:54 IL30_4971:1:52:15193:6126/2 26 76 + Contig_000301929 10540 10590 + 48 M 51
cigar:A:53 IL30_4971:1:72:10979:17000/1 1 76 + Contig_000301929 10543 10618 + 76 M 76
cigar:A:53 IL30_4971:1:6:3063:8782/1 76 19 - Contig_000301929 10556 10613 + 52 M 58
cigar:A:52 IL30_4971:1:117:16235:15376/1 32 76 + Contig_000301929 10585 10629 + 45 M 45
cigar:A:54 IL30_4971:1:119:16773:5373/1 1 76 + Contig_000301929 10596 10671 + 73 M 76
cigar:A:53 IL30_4971:1:111:5678:17224/1 1 76 + Contig_000301929 10608 10683 + 76 M 76
cigar:A:52 IL30_4971:1:103:14341:20162/1 76 1 - Contig_000301929 10651 10726 + 76 M 76
cigar:A:51 IL30_4971:1:18:14105:4765/1 76 1 - Contig_000301929 10654 10729 + 76 M 76
cigar:A:45 IL30_4971:1:77:11727:10964/1 76 20 - Contig_000301929 10669 10725 + 57 M 57
cigar:A:54 IL30_4971:1:117:14103:10528/2 76 1 - Contig_000301929 10720 10795 + 69 M 76
cigar:A:57 IL30_4971:1:21:16737:3484/2 27 76 + Contig_000301929 10770 10819 + 50 M 50
cigar:A:56 IL30_4971:1:77:17727:1889/2 76 1 - Contig_000301929 10773 10848 + 66 M 76
cigar:A:50 IL30_4971:1:14:13832:3494/1 24 76 + Contig_000301929 10809 10861 + 50 M 53
cigar:A:56 IL30_4971:1:81:10641:5816/1 75 1 - Contig_000301929 10809 10884 + 41 M 10 I 1 M 10 D 1 M 25 D 1 M 29
cigar:A:47 IL30_4971:1:51:6177:21038/2 1 76 + Contig_000301929 10832 10908 + 65 M 43 D 1 M 33
cigar:A:49 IL30_4971:1:37:3799:3844/1 1 76 + Contig_000301929 10851 10927 + 72 M 24 D 1 M 52
cigar:A:52 IL30_4971:1:117:16235:15376/2 76 1 - Contig_000301929 10986 11061 + 66 M 76
cigar:A:54 IL30_4971:1:52:15193:6126/1 76 1 - Contig_000301929 10998 11073 + 76 M 76
cigar:A:53 IL30_4971:1:72:10979:17000/2 76 45 - Contig_000301929 10999 11030 + 32 M 32
cigar:A:53 IL30_4971:1:84:11619:8261/2 76 1 - Contig_000301929 11010 11085 + 67 M 76
cigar:A:50 IL30_4971:1:63:12456:14789/2 1 76 + Contig_000301929 11041 11116 + 69 M 76
cigar:A:54 IL30_4971:1:119:16773:5373/2 76 39 - Contig_000301929 11049 11086 + 38 M 38
cigar:A:53 IL30_4971:1:111:5678:17224/2 76 40 - Contig_000301929 11050 11086 + 37 M 37
cigar:A:51 IL30_4971:1:19:3295:16268/1 22 76 + Contig_000301929 11107 11161 + 55 M 55
cigar:A:53 IL30_4971:1:51:11180:15151/2 21 76 + Contig_000301929 11107 11162 + 49 M 56
cigar:A:54 IL30_4971:1:59:19062:2913/2 26 76 + Contig_000301929 11136 11186 + 51 M 51
cigar:A:52 IL30_4971:1:31:10000:16650/2 9 76 + Contig_000301929 11139 11206 + 60 M 68
cigar:A:52 IL30_4971:1:81:14536:13550/2 9 76 + Contig_000301929 11139 11206 + 63 M 68
cigar:A:57 IL30_4971:1:21:16737:3484/1 76 1 - Contig_000301929 11169 11244 + 76 M 76
cigar:A:50 IL30_4971:1:72:14197:13505/2 28 76 + Contig_000301929 11263 11311 + 49 M 49
cigar:A:50 IL30_4971:1:14:13832:3494/2 76 1 - Contig_000301929 11311 11386 + 69 M 76
cigar:A:49 IL30_4971:1:37:3799:3844/2 76 40 - Contig_000301929 11330 11366 + 34 M 37
cigar:A:52 IL30_4971:1:84:3957:3792/2 1 76 + Contig_000301929 11363 11438 + 70 M 76
cigar:A:47 IL30_4971:1:51:6177:21038/1 76 15 - Contig_000301929 11374 11435 + 53 M 62
cigar:A:52 IL30_4971:1:81:14536:13550/1 76 22 - Contig_000301929 11398 11452 + 49 M 55
cigar:A:52 IL30_4971:1:31:10000:16650/1 75 22 - Contig_000301929 11399 11452 + 48 M 54
cigar:A:50 IL30_4971:1:72:14197:13505/1 76 1 - Contig_000301929 11493 11568 + 76 M 76
cigar:A:50 IL30_4971:1:63:12456:14789/1 76 38 - Contig_000301929 11494 11532 + 39 M 39
cigar:A:51 IL30_4971:1:19:3295:16268/2 76 1 - Contig_000301929 11556 11631 + 66 M 76
cigar:A:53 IL30_4971:1:51:11180:15151/1 76 22 - Contig_000301929 11574 11628 + 55 M 55
cigar:A:54 IL30_4971:1:59:19062:2913/1 76 1 - Contig_000301929 11597 11672 + 73 M 76
cigar:A:52 IL30_4971:1:84:3957:3792/1 76 24 - Contig_000301929 11787 11839 + 53 M 53
(END)
```

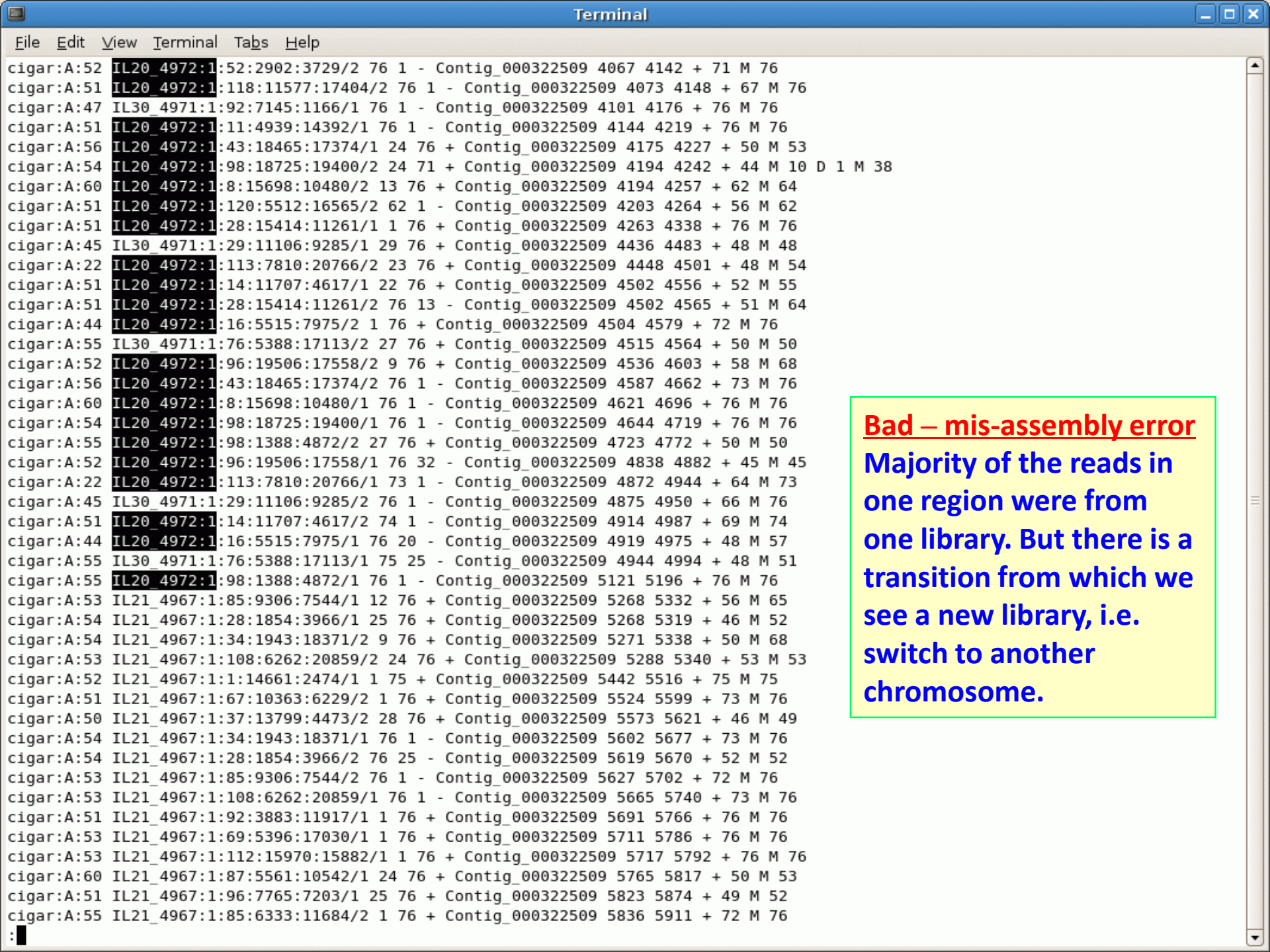
Perfect - Reads from the same library were mapped to the contig

Terminal

File Edit View Terminal Tabs Help

```
cigar:A:27 IL20_4972:1:34:5929:21194/2 1 76 + Contig_000284484 69 144 + 73 M 76
cigar:A:60 IL20_4972:1:98:3543:2996/2 1 76 + Contig_000284484 135 210 + 71 M 76
cigar:A:58 IL30_4971:1:55:12151:7461/2 24 76 + Contig_000284484 163 215 + 53 M 53
cigar:A:58 IL30_4971:1:55:12156:7489/2 24 76 + Contig_000284484 163 215 + 53 M 53
cigar:A:45 IL20_4972:1:80:11636:19064/2 23 76 + Contig_000284484 305 358 + 51 M 54
cigar:A:14 IL20_4972:1:63:6677:17740/1 17 76 + Contig_000284484 308 367 + 51 M 60
cigar:A:60 IL20_4972:1:95:15961:4421/2 1 76 + Contig_000284484 410 485 + 68 M 76
cigar:A:17 IL20_4972:1:22:16269:8052/2 32 74 + Contig_000284484 452 494 + 40 M 43
cigar:A:27 IL20_4972:1:34:5929:21194/1 76 25 - Contig_000284484 467 518 + 49 M 52
cigar:A:33 IL20_4972:1:35:7687:21118/2 1 73 + Contig_000284484 494 566 + 70 M 73
cigar:A:60 IL20_4972:1:98:3543:2996/1 76 1 - Contig_000284484 568 643 + 76 M 76
cigar:A:58 IL30_4971:1:55:12151:7461/1 76 1 - Contig_000284484 592 667 + 76 M 76
cigar:A:58 IL30_4971:1:55:12156:7489/1 76 1 - Contig_000284484 592 667 + 76 M 76
cigar:A:60 IL20_4972:1:7:10184:12209/2 15 76 + Contig_000284484 618 679 + 62 M 62
cigar:A:45 IL20_4972:1:80:11636:19064/1 76 1 - Contig_000284484 724 799 + 75 M 76
cigar:A:14 IL20_4972:1:63:6677:17740/2 73 1 - Contig_000284484 732 804 + 66 M 73
cigar:A:17 IL20_4972:1:22:16269:8052/1 76 1 - Contig_000284484 870 945 + 76 M 76
cigar:A:33 IL20_4972:1:35:7687:21118/1 76 1 - Contig_000284484 882 957 + 72 M 76
cigar:A:60 IL20_4972:1:95:15961:4421/1 76 1 - Contig_000284484 893 968 + 76 M 76
cigar:A:52 IL20_4972:1:42:14253:8496/1 26 76 + Contig_000284484 961 1011 + 51 M 51
cigar:A:19 IL20_4972:1:50:18127:6058/2 8 72 + Contig_000284484 967 1032 + 52 M 11 D 1 M 54
cigar:A:51 IL20_4972:1:66:7775:15830/1 14 76 + Contig_000284484 971 1033 + 63 M 63
cigar:A:60 IL20_4972:1:24:19710:11776/1 12 76 + Contig_000284484 971 1035 + 65 M 65
cigar:A:60 IL30_4971:1:52:1801:12531/1 14 76 + Contig_000284484 971 1033 + 63 M 63
cigar:A:60 IL30_4971:1:52:1811:12546/1 14 76 + Contig_000284484 971 1033 + 63 M 63
cigar:A:60 IL20_4972:1:7:10184:12209/1 76 1 - Contig_000284484 1006 1081 + 76 M 76
cigar:A:12 IL30_4971:1:73:16275:14156/2 27 73 + Contig_000284484 1228 1274 + 47 M 47
cigar:A:52 IL20_4972:1:42:14253:8496/2 74 1 - Contig_000284484 1334 1407 + 71 M 74
cigar:A:51 IL20_4972:1:66:7775:15830/2 76 1 - Contig_000284484 1359 1434 + 72 M 76
cigar:A:60 IL20_4972:1:24:19710:11776/2 76 1 - Contig_000284484 1365 1440 + 73 M 76
cigar:A:13 IL20_4972:1:89:8092:8635/1 21 76 + Contig_000284484 1368 1423 + 56 M 56
cigar:A:50 IL20_4972:1:41:19227:21028/1 23 76 + Contig_000284484 1428 1481 + 48 M 54
cigar:A:19 IL20_4972:1:50:18127:6058/1 76 1 - Contig_000284484 1429 1504 + 76 M 76
cigar:A:12 IL30_4971:1:73:16275:14156/1 76 1 - Contig_000284484 1452 1527 + 76 M 76
cigar:A:60 IL30_4971:1:52:1801:12531/2 76 1 - Contig_000284484 1456 1531 + 68 M 76
cigar:A:60 IL30_4971:1:52:1811:12546/2 76 1 - Contig_000284484 1456 1531 + 68 M 76
cigar:A:52 IL20_4972:1:102:7054:7616/1 22 76 + Contig_000284484 1519 1573 + 55 M 55
cigar:A:26 IL30_4971:1:104:5240:4338/1 23 76 + Contig_000284484 1522 1575 + 51 M 54
cigar:A:39 IL20_4972:1:61:11878:2929/2 25 76 + Contig_000284484 1600 1651 + 49 M 52
cigar:A:60 IL20_4972:1:83:12689:14515/1 21 76 + Contig_000284484 1710 1765 + 53 M 56
cigar:A:60 IL20_4972:1:27:8709:17100/2 26 76 + Contig_000284484 1765 1815 + 48 M 51
cigar:A:13 IL20_4972:1:89:8092:8635/2 71 28 - Contig_000284484 1794 1837 + 44 M 44
cigar:A:38 IL30_4971:1:103:13980:15445/2 26 76 + Contig_000284484 1823 1873 + 51 M 51
cigar:A:51 IL30_4971:1:95:8853:8709/2 21 76 + Contig_000284484 1824 1879 + 44 M 56
:
```

Acceptable - Majority of the reads were from the same library, but there were reads from other libraries


Terminal
File Edit View Terminal Tabs Help
cigar:A:52 IL20_4972:1:52:2902:3729/2 76 1 - Contig_000322509 4067 4142 + 71 M 76
cigar:A:51 IL20_4972:1:118:11577:17404/2 76 1 - Contig_000322509 4073 4148 + 67 M 76
cigar:A:47 IL30_4971:1:92:7145:1166/1 76 1 - Contig_000322509 4101 4176 + 76 M 76
cigar:A:51 IL20_4972:1:11:4939:14392/1 76 1 - Contig_000322509 4144 4219 + 76 M 76
cigar:A:56 IL20_4972:1:43:18465:17374/1 24 76 + Contig_000322509 4175 4227 + 50 M 53
cigar:A:54 IL20_4972:1:98:18725:19400/2 24 71 + Contig_000322509 4194 4242 + 44 M 10 D 1 M 38
cigar:A:60 IL20_4972:1:8:15698:10480/2 13 76 + Contig_000322509 4194 4257 + 62 M 64
cigar:A:51 IL20_4972:1:120:5512:16565/2 62 1 - Contig_000322509 4203 4264 + 56 M 62
cigar:A:51 IL20_4972:1:28:15414:11261/1 1 76 + Contig_000322509 4263 4338 + 76 M 76
cigar:A:45 IL30_4971:1:29:11106:9285/1 29 76 + Contig_000322509 4436 4483 + 48 M 48
cigar:A:22 IL20_4972:1:113:7810:20766/2 23 76 + Contig_000322509 4448 4501 + 48 M 54
cigar:A:51 IL20_4972:1:14:11707:4617/1 22 76 + Contig_000322509 4502 4556 + 52 M 55
cigar:A:51 IL20_4972:1:28:15414:11261/2 76 13 - Contig_000322509 4502 4565 + 51 M 64
cigar:A:44 IL20_4972:1:16:5515:7975/2 1 76 + Contig_000322509 4504 4579 + 72 M 76
cigar:A:55 IL30_4971:1:76:5388:17113/2 27 76 + Contig_000322509 4515 4564 + 50 M 50
cigar:A:52 IL20_4972:1:96:19506:17558/2 9 76 + Contig_000322509 4536 4603 + 58 M 68
cigar:A:56 IL20_4972:1:43:18465:17374/2 76 1 - Contig_000322509 4587 4662 + 73 M 76
cigar:A:60 IL20_4972:1:8:15698:10480/1 76 1 - Contig_000322509 4621 4696 + 76 M 76
cigar:A:54 IL20_4972:1:98:18725:19400/1 76 1 - Contig_000322509 4644 4719 + 76 M 76
cigar:A:55 IL20_4972:1:98:1388:4872/2 27 76 + Contig_000322509 4723 4772 + 50 M 50
cigar:A:52 IL20_4972:1:96:19506:17558/1 76 32 - Contig_000322509 4838 4882 + 45 M 45
cigar:A:22 IL20_4972:1:113:7810:20766/1 73 1 - Contig_000322509 4872 4944 + 64 M 73
cigar:A:45 IL30_4971:1:29:11106:9285/2 76 1 - Contig_000322509 4875 4950 + 66 M 76
cigar:A:51 IL20_4972:1:14:11707:4617/2 74 1 - Contig_000322509 4914 4987 + 69 M 74
cigar:A:44 IL20_4972:1:16:5515:7975/1 76 20 - Contig_000322509 4919 4975 + 48 M 57
cigar:A:55 IL30_4971:1:76:5388:17113/1 75 25 - Contig_000322509 4944 4994 + 48 M 51
cigar:A:55 IL20_4972:1:98:1388:4872/1 76 1 - Contig_000322509 5121 5196 + 76 M 76
cigar:A:53 IL21_4967:1:85:9306:7544/1 12 76 + Contig_000322509 5268 5332 + 56 M 65
cigar:A:54 IL21_4967:1:28:1854:3966/1 25 76 + Contig_000322509 5268 5319 + 46 M 52
cigar:A:54 IL21_4967:1:34:1943:18371/2 9 76 + Contig_000322509 5271 5338 + 50 M 68
cigar:A:53 IL21_4967:1:108:6262:20859/2 24 76 + Contig_000322509 5288 5340 + 53 M 53
cigar:A:52 IL21_4967:1:1:14661:2474/1 1 75 + Contig_000322509 5442 5516 + 75 M 75
cigar:A:51 IL21_4967:1:67:10363:6229/2 1 76 + Contig_000322509 5524 5599 + 73 M 76
cigar:A:50 IL21_4967:1:37:13799:4473/2 28 76 + Contig_000322509 5573 5621 + 46 M 49
cigar:A:54 IL21_4967:1:34:1943:18371/1 76 1 - Contig_000322509 5602 5677 + 73 M 76
cigar:A:54 IL21_4967:1:28:1854:3966/2 76 25 - Contig_000322509 5619 5670 + 52 M 52
cigar:A:53 IL21_4967:1:85:9306:7544/2 76 1 - Contig_000322509 5627 5702 + 72 M 76
cigar:A:53 IL21_4967:1:108:6262:20859/1 76 1 - Contig_000322509 5665 5740 + 73 M 76
cigar:A:51 IL21_4967:1:92:3883:11917/1 1 76 + Contig_000322509 5691 5766 + 76 M 76
cigar:A:53 IL21_4967:1:69:5396:17030/1 1 76 + Contig_000322509 5711 5786 + 76 M 76
cigar:A:53 IL21_4967:1:112:15970:15882/1 1 76 + Contig_000322509 5717 5792 + 76 M 76
cigar:A:60 IL21_4967:1:87:5561:10542/1 24 76 + Contig_000322509 5765 5817 + 50 M 53
cigar:A:51 IL21_4967:1:96:7765:7203/1 25 76 + Contig_000322509 5823 5874 + 49 M 52
cigar:A:55 IL21_4967:1:85:6333:11684/2 1 76 + Contig_000322509 5836 5911 + 72 M 76
:
Bad – mis-assembly error
Majority of the reads in one region were from one library. But there is a transition from which we see a new library, i.e. switch to another chromosome.

Terminal

File Edit View Terminal Tabs Help

```
supercontig Chr1_contigs_000012321
contig Chr1_contigs_000012321
gap 500 * * *
contig Chr1_contigs_000058081
gap 500 * * *
contig Chr1_contigs_000008504
gap 500 * * *
contig Unassign_contigs_000000011
gap 500 * * *
contig Chr1_contigs_000053754
gap 500 * * *
contig Chr1_contigs_000042990
gap 500 * * *
contig Chr1_contigs_000030832
gap 500 * * *
contig Chr1_contigs_000006882
gap 500 * * *
contig Chr1_contigs_000006880
gap 500 * * *
contig Chr1_contigs_000016870
gap 500 * * *
contig Chr1_contigs_000013219
gap 500 * * *
contig Chr1_contigs_000013921

supercontig Chr1_contigs_000020966
contig Chr1_contigs_000020966
gap 500 * * *
contig Chr1_contigs_000056393
gap 500 * * *
contig Chr1_contigs_000047464
gap 500 * * *
contig Chr1_contigs_000057808
gap 500 * * *
contig Chr1_contigs_000006519
gap 500 * * *
:
```

Unassigned contigs were placed by supercontigs using mate pairs

Scaffolds Assigned to Chromosomes using Flow-sorting Data

Chr_ID	Chr_size	Scaffolds_assigned	Bases_assigned Mb
Chr1	571	6729	684
Chr2	610	8381	740
Chr3	556	7197	641
Chr4	450	4817	487
Chr5	341	3188	300
Chr6	277	2844	263
Chrx	122	2378	86.6
Unassigned		440	1.23

Genome Assembly Normal – T. Devil

Solexa reads:

Number of read pairs:	650 Million;
Finished genome size:	3.3 GB;
Read length:	2x100bp;
Estimated read coverage:	~40X;
Insert size:	410/50-600 bp;
Mate pair data:	2k,4k,5k,6k,8k,10k
Number of reads clustered:	591 Million

Assembly features: - stats

	Contigs	Supercontigs
Total number of contigs:	237,291	35,974
Total bases of contigs:	2.93 Gb	3.17 Gb
N50 contig size:	20,139	1,847,186
Largest contig:	189,866	5,315,556
Averaged contig size:	12,354	88,254
Contig coverage on genome:	~94%	>99%
Ratio of placed PE reads:	~92%	?

Devil Tumour Genome Assemblies

Solexa reads:	Tumour_87T	Tumour_53T
Number of read pairs:	760 Million	669 M;
Finished genome size:	3.3 GB	3.3 GB;
Read length:	2x100	2x100;
Estimated read coverage:	~46X	~40X;
Insert size:	300bp	300bp;
Number of reads clustered:	635 Million	603 M

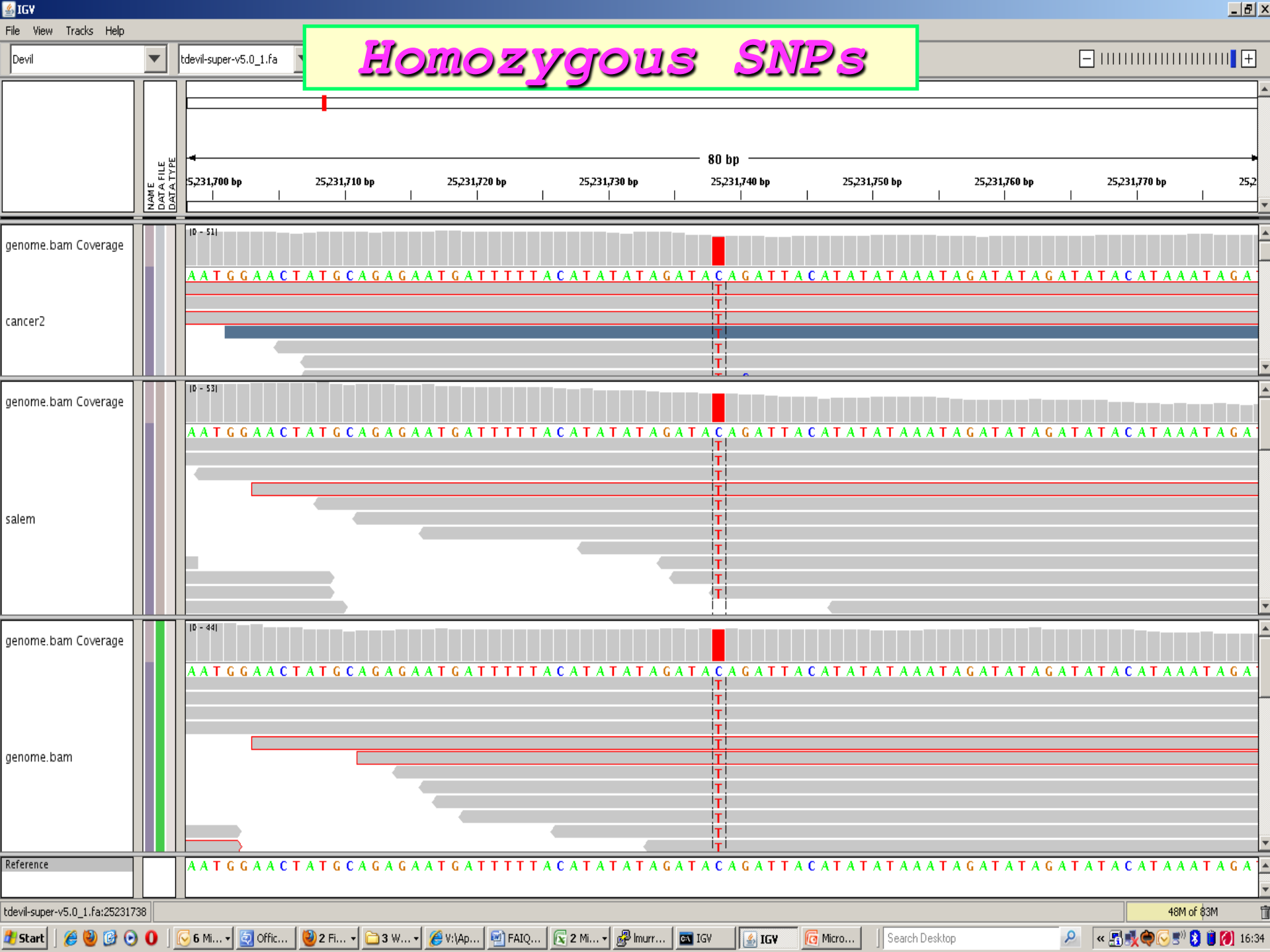
Assembly features: - stats

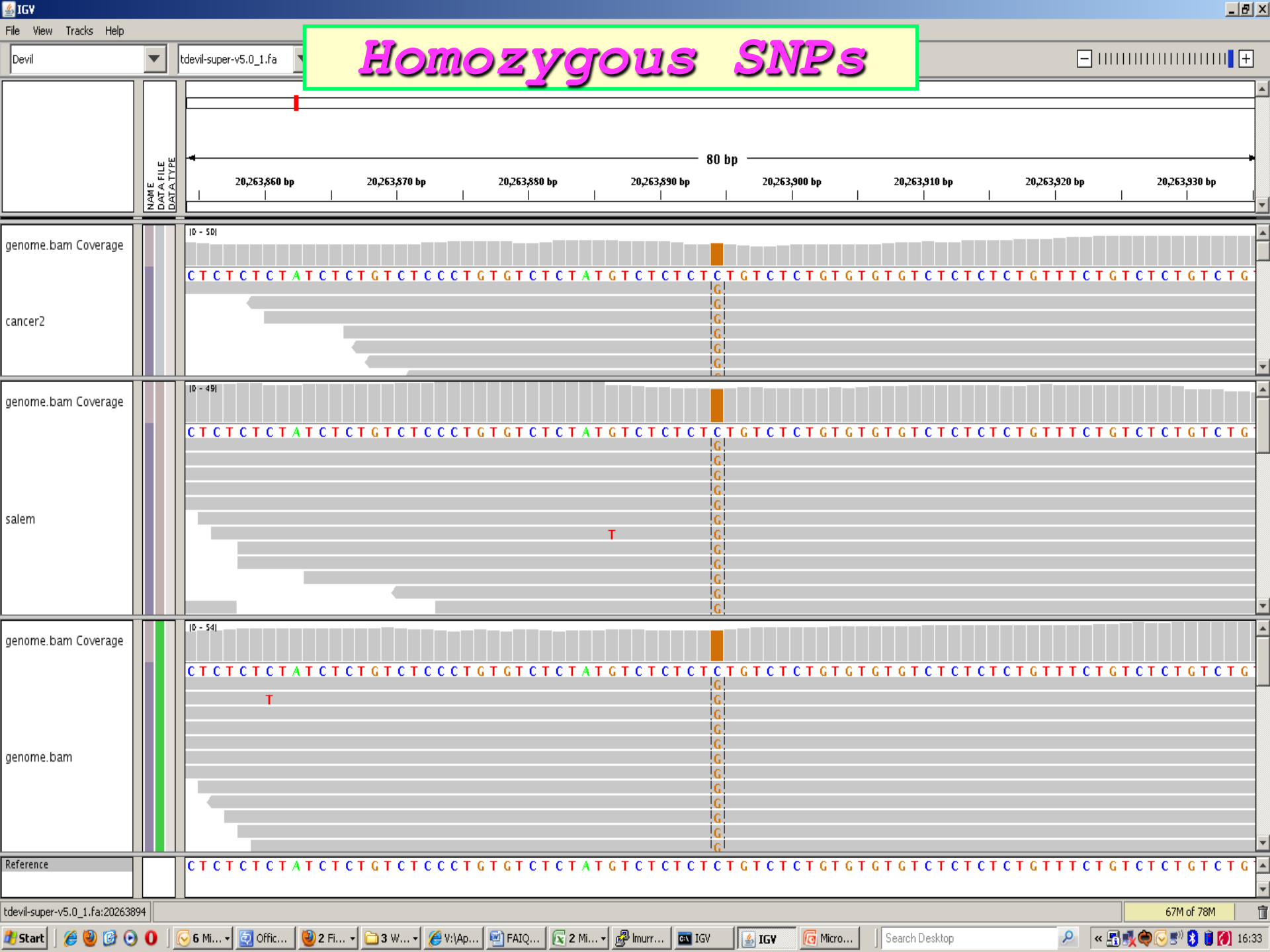
	Tumour_87T	Tumour_53T
Total number of contigs:	532,584	612,288
Total bases of contigs:	3.13 Gb	3.14 Gb
N50 contig size:	15,908	14,632
Largest contig:	109,065	170,831
Averaged contig size:	5,882	5,567
Contig coverage on genome:	~95%	~95%
Ratio of placed PE reads:	~92%	~92%

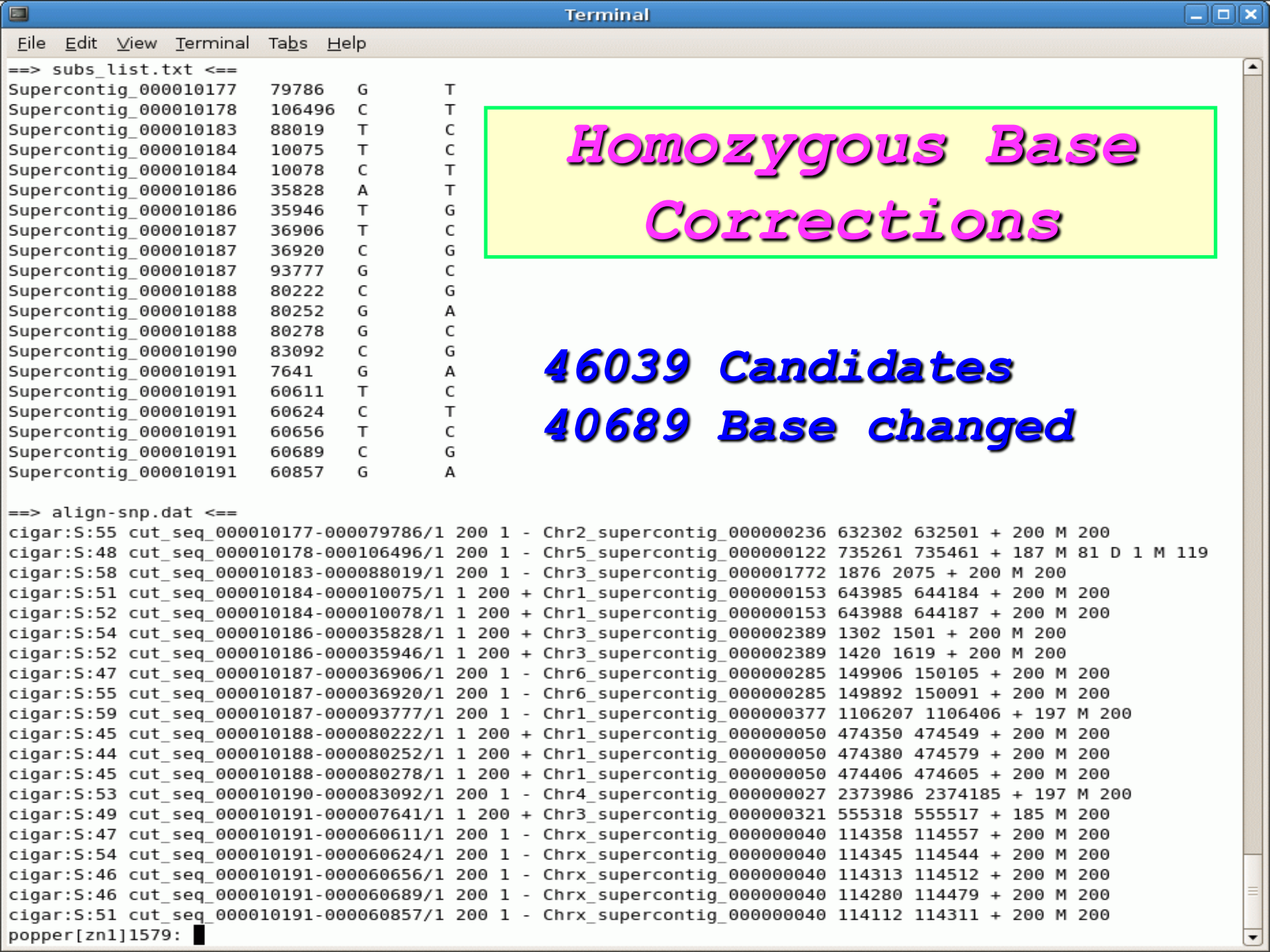
Variant calling : catalogue of variants in all 4 genomes

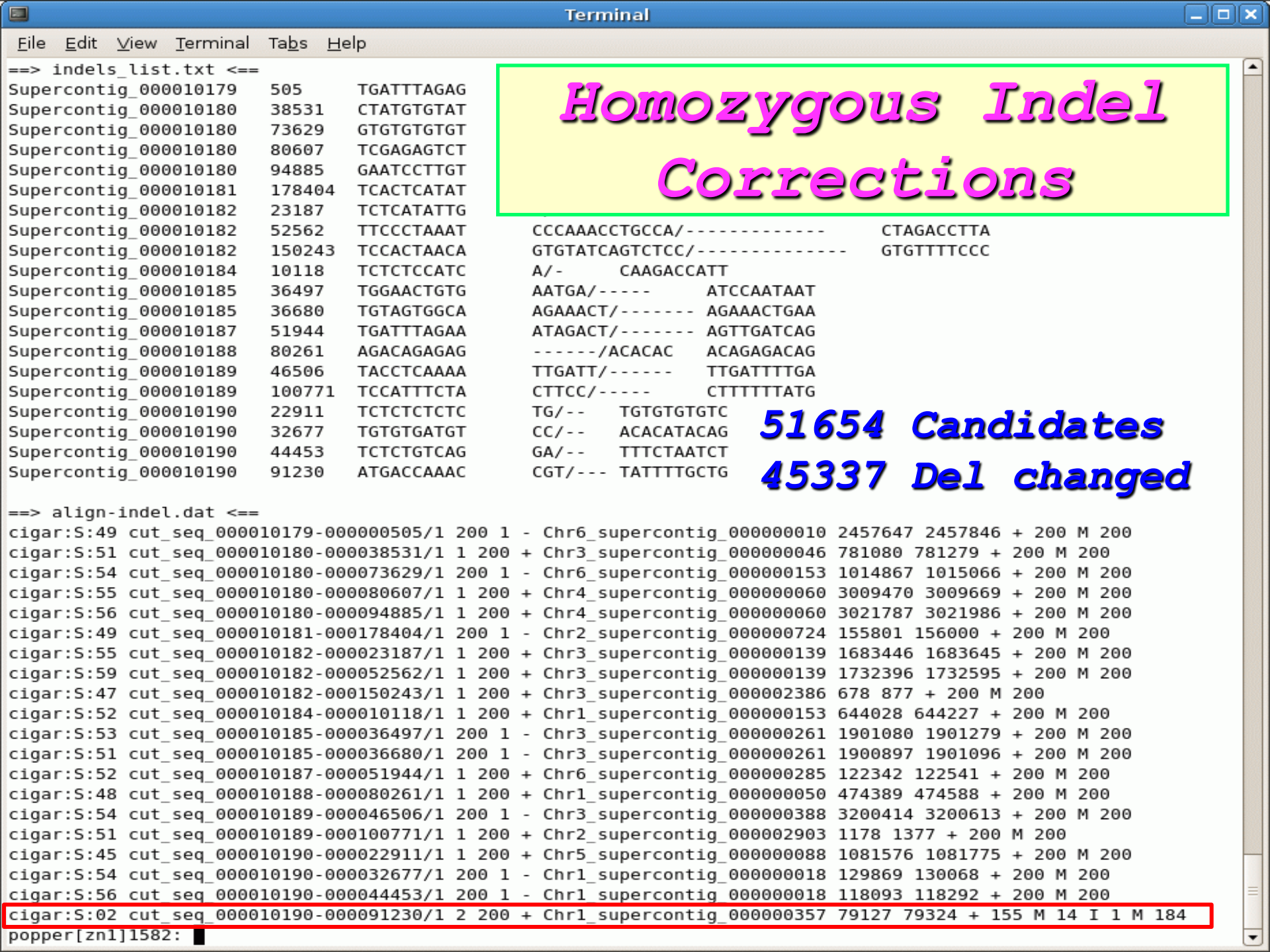
	Salem (91H)	Joey (31H)	Cancer 1 (87T)	Cancer 2 (53T)
Coverage	35.58	28.80	40.49	33.14
Total SNPs	615,084	646,186	758,023	738,793
Het SNPs	524,040	371,412	465,630	462,722
Hom SNPs	91,044	274,774	292,393	276,071
Total indels	235,632	262,461	320,820	312,287
Het indels	183,978	146,299	186,094	183,747
Hom indels	51,654	81,120 / 116,162	134,726	128,540

***Data source: Illumina. Variants removed within 500bp of a contig end, Q(indel) < 30 and Q(GT) < 5.**

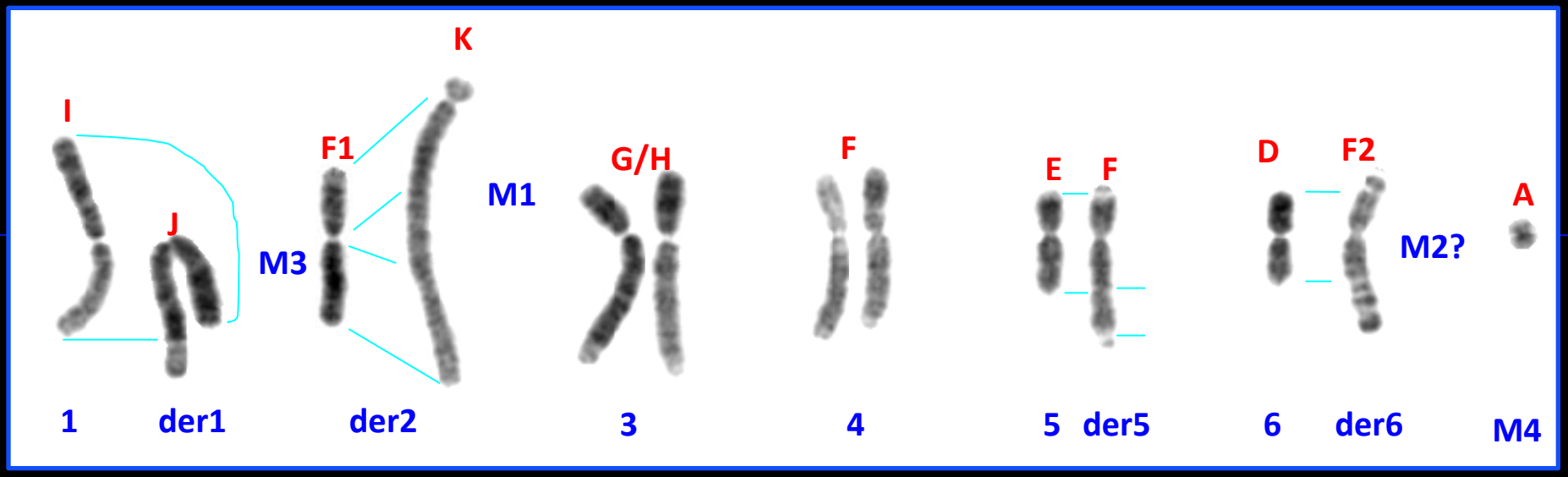

IGV
File View Tracks Help
Devil
tdevil-super-v5.0_1.fa
Homozygous SNPs
NAME
DATA FILE
DATA TYPE
80 bp
25,231,700 bp
25,231,710 bp
25,231,720 bp
25,231,730 bp
25,231,740 bp
25,231,750 bp
25,231,760 bp
25,231,770 bp
genome.bam Coverage
cancer2
genome.bam Coverage
salem
genome.bam Coverage
genome.bam
Reference
A A T G G A A C T A T G C A G A G A A T G A T T T T T A C A T A T A T A G A T A C A G A T T A C A T A T A T A A A T A G A T A T A G A T A T A C A T A A A T A G A
tdevil-super-v5.0_1.fa:25231738
48M of 83M
Start
Search Desktop
16:34

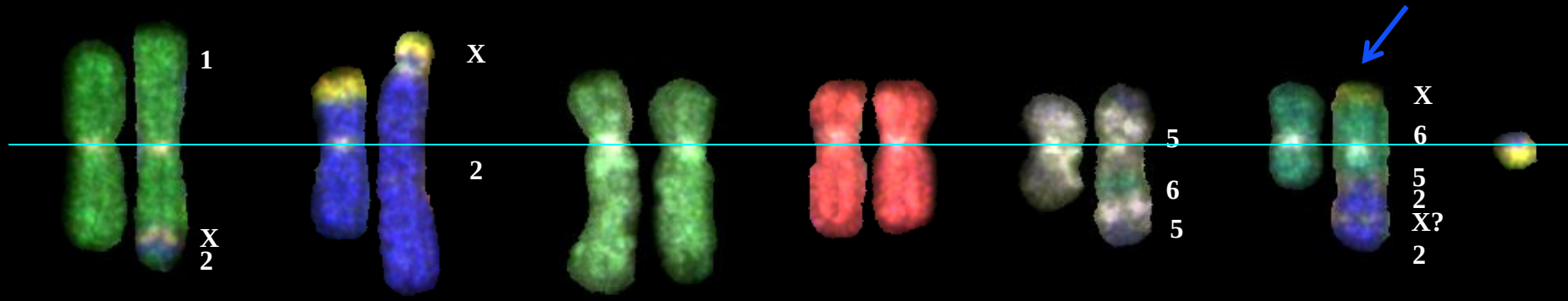

IGV
File View Tracks Help
Devil
tdevil-super-v5.0_1.fa
Homozygous SNPs
NAME
DATA FILE
DATA TYPE
80 bp
20,263,860 bp
20,263,870 bp
20,263,880 bp
20,263,890 bp
20,263,900 bp
20,263,910 bp
20,263,920 bp
20,263,930 bp
genome.bam Coverage
cancer2
genome.bam Coverage
salem
genome.bam Coverage
genome.bam
Reference
tdevil-super-v5.0_1.fa:20263894
67M of 78M
Start
Search Desktop
16:33


Terminal
File Edit View Terminal Tabs Help
==> subs_list.txt <==
Supercontig_000010177 79786 G T
Supercontig_000010178 106496 C T
Supercontig_000010183 88019 T C
Supercontig_000010184 10075 T C
Supercontig_000010184 10078 C T
Supercontig_000010186 35828 A T
Supercontig_000010186 35946 T G
Supercontig_000010187 36906 T C
Supercontig_000010187 36920 C G
Supercontig_000010187 93777 G C
Supercontig_000010188 80222 C G
Supercontig_000010188 80252 G A
Supercontig_000010188 80278 G C
Supercontig_000010190 83092 C G
Supercontig_000010191 7641 G A
Supercontig_000010191 60611 T C
Supercontig_000010191 60624 C T
Supercontig_000010191 60656 T C
Supercontig_000010191 60689 C G
Supercontig_000010191 60857 G A
Homozygous Base Corrections
46039 Candidates
40689 Base changed
==> align-snp.dat <==
cigar:S:55 cut_seq_000010177-000079786/1 200 1 - Chr2_supercontig_000000236 632302 632501 + 200 M 200
cigar:S:48 cut_seq_000010178-000106496/1 200 1 - Chr5_supercontig_000000122 735261 735461 + 187 M 81 D 1 M 119
cigar:S:58 cut_seq_000010183-000088019/1 200 1 - Chr3_supercontig_000001772 1876 2075 + 200 M 200
cigar:S:51 cut_seq_000010184-000010075/1 1 200 + Chr1_supercontig_000000153 643985 644184 + 200 M 200
cigar:S:52 cut_seq_000010184-000010078/1 1 200 + Chr1_supercontig_000000153 643988 644187 + 200 M 200
cigar:S:54 cut_seq_000010186-000035828/1 1 200 + Chr3_supercontig_000002389 1302 1501 + 200 M 200
cigar:S:52 cut_seq_000010186-000035946/1 1 200 + Chr3_supercontig_000002389 1420 1619 + 200 M 200
cigar:S:47 cut_seq_000010187-000036906/1 200 1 - Chr6_supercontig_000000285 149906 150105 + 200 M 200
cigar:S:55 cut_seq_000010187-000036920/1 200 1 - Chr6_supercontig_000000285 149892 150091 + 200 M 200
cigar:S:59 cut_seq_000010187-000093777/1 200 1 - Chr1_supercontig_000000377 1106207 1106406 + 197 M 200
cigar:S:45 cut_seq_000010188-000080222/1 1 200 + Chr1_supercontig_000000050 474350 474549 + 200 M 200
cigar:S:44 cut_seq_000010188-000080252/1 1 200 + Chr1_supercontig_000000050 474380 474579 + 200 M 200
cigar:S:45 cut_seq_000010188-000080278/1 1 200 + Chr1_supercontig_000000050 474406 474605 + 200 M 200
cigar:S:53 cut_seq_000010190-000083092/1 200 1 - Chr4_supercontig_000000027 2373986 2374185 + 197 M 200
cigar:S:49 cut_seq_000010191-000007641/1 1 200 + Chr3_supercontig_000000321 555318 555517 + 185 M 200
cigar:S:47 cut_seq_000010191-000060611/1 200 1 - Chrx_supercontig_000000040 114358 114557 + 200 M 200
cigar:S:54 cut_seq_000010191-000060624/1 200 1 - Chrx_supercontig_000000040 114345 114544 + 200 M 200
cigar:S:46 cut_seq_000010191-000060656/1 200 1 - Chrx_supercontig_000000040 114313 114512 + 200 M 200
cigar:S:46 cut_seq_000010191-000060689/1 200 1 - Chrx_supercontig_000000040 114280 114479 + 200 M 200
cigar:S:51 cut_seq_000010191-000060857/1 200 1 - Chrx_supercontig_000000040 114112 114311 + 200 M 200
popper[zn1]1579:

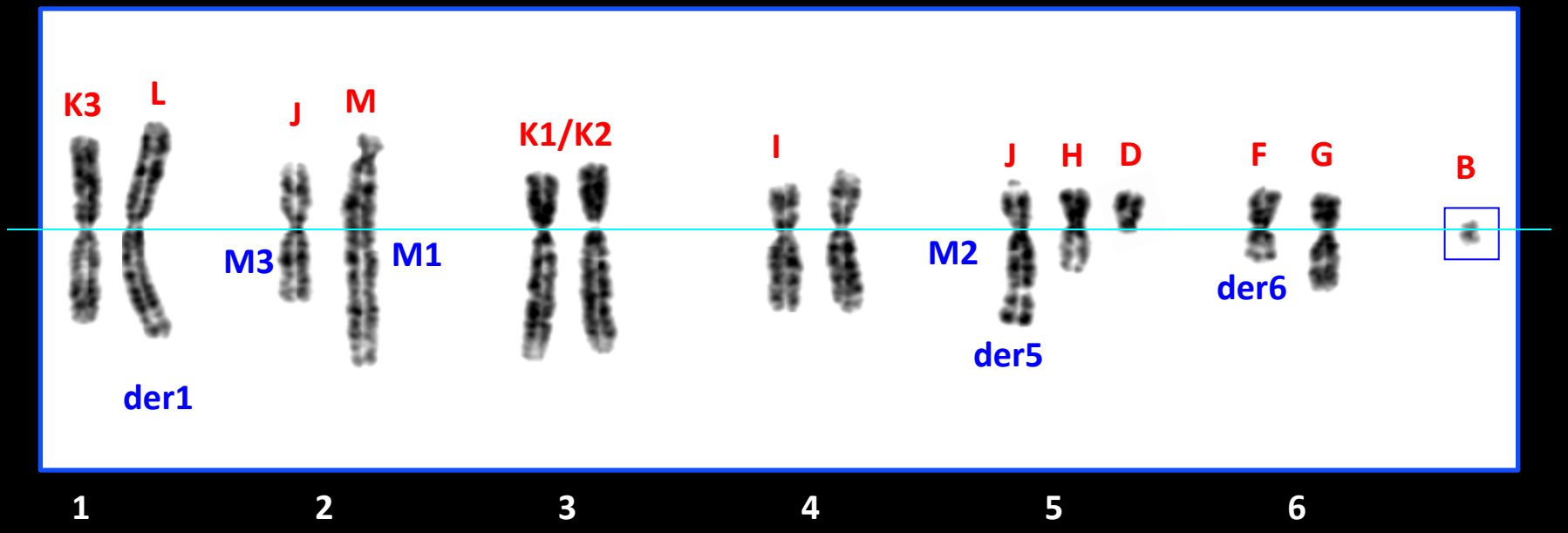

Terminal
File Edit View Terminal Tabs Help
==> indels_list.txt <==
Supercontig_000010179 505 TGATTTAGAG
Supercontig_000010180 38531 CTATGTGTAT
Supercontig_000010180 73629 GTGTGTGTGT
Supercontig_000010180 80607 TCGAGAGTCT
Supercontig_000010180 94885 GAATCCTTGT
Supercontig_000010181 178404 TCACTCATAT
Supercontig_000010182 23187 TCTCATATTG
Supercontig_000010182 52562 TTCCCTAAAT CCCAAACCTGCCA/---------- CTAGACCTTA
Supercontig_000010182 150243 TCCACTAACA GTGTATCAGTCTCC/---------- GTGTTTTCCC
Supercontig_000010184 10118 TCTCTCCATC A/- CAAGACCATT
Supercontig_000010185 36497 TGGAACTGTG AATGA/----- ATCCAATAAT
Supercontig_000010185 36680 TGTAGTGGCA AGAAACT/------- AGAAACTGAA
Supercontig_000010187 51944 TGATTTAGAA ATAGACT/------- AGTTGATCAG
Supercontig_000010188 80261 AGACAGAGAG ------/ACACAC ACAGAGACAG
Supercontig_000010189 46506 TACCTCAAAA TTGATT/------ TTGATTTTGA
Supercontig_000010189 100771 TCCATTTCTA CTTCC/----- CTTTTTTATG
Supercontig_000010190 22911 TCTCTCTCTC TG/-- TGTGTGTGTC
Supercontig_000010190 32677 TGTGTGATGT CC/-- ACACATACAG
Supercontig_000010190 44453 TCTCTGTCAG GA/-- TTTCTAATCT
Supercontig_000010190 91230 ATGACCAAAC CGT/--- TATTTTGCTG
Homozygous Indel Corrections
51654 Candidates
45337 Del changed
==> align-indel.dat <==
cigar:S:49 cut_seq_000010179-000000505/1 200 1 - Chr6_supercontig_000000010 2457647 2457846 + 200 M 200
cigar:S:51 cut_seq_000010180-000038531/1 1 200 + Chr3_supercontig_000000046 781080 781279 + 200 M 200
cigar:S:54 cut_seq_000010180-000073629/1 200 1 - Chr6_supercontig_000000153 1014867 1015066 + 200 M 200
cigar:S:55 cut_seq_000010180-000080607/1 1 200 + Chr4_supercontig_000000060 3009470 3009669 + 200 M 200
cigar:S:56 cut_seq_000010180-000094885/1 1 200 + Chr4_supercontig_000000060 3021787 3021986 + 200 M 200
cigar:S:49 cut_seq_000010181-000178404/1 200 1 - Chr2_supercontig_000000724 155801 156000 + 200 M 200
cigar:S:55 cut_seq_000010182-000023187/1 1 200 + Chr3_supercontig_000000139 1683446 1683645 + 200 M 200
cigar:S:59 cut_seq_000010182-000052562/1 1 200 + Chr3_supercontig_000000139 1732396 1732595 + 200 M 200
cigar:S:47 cut_seq_000010182-000150243/1 1 200 + Chr3_supercontig_000002386 678 877 + 200 M 200
cigar:S:52 cut_seq_000010184-000010118/1 1 200 + Chr1_supercontig_000000153 644028 644227 + 200 M 200
cigar:S:53 cut_seq_000010185-000036497/1 200 1 - Chr3_supercontig_000000261 1901080 1901279 + 200 M 200
cigar:S:51 cut_seq_000010185-000036680/1 200 1 - Chr3_supercontig_000000261 1900897 1901096 + 200 M 200
cigar:S:52 cut_seq_000010187-000051944/1 1 200 + Chr6_supercontig_000000285 122342 122541 + 200 M 200
cigar:S:48 cut_seq_000010188-000080261/1 1 200 + Chr1_supercontig_000000050 474389 474588 + 200 M 200
cigar:S:54 cut_seq_000010189-000046506/1 200 1 - Chr3_supercontig_000000388 3200414 3200613 + 200 M 200
cigar:S:51 cut_seq_000010189-000100771/1 1 200 + Chr2_supercontig_000002903 1178 1377 + 200 M 200
cigar:S:45 cut_seq_000010190-000022911/1 1 200 + Chr5_supercontig_000000088 1081576 1081775 + 200 M 200
cigar:S:54 cut_seq_000010190-000032677/1 200 1 - Chr1_supercontig_000000018 129869 130068 + 200 M 200
cigar:S:56 cut_seq_000010190-000044453/1 200 1 - Chr1_supercontig_000000018 118093 118292 + 200 M 200
cigar:S:02 cut_seq_000010190-000091230/1 2 200 + Chr1_supercontig_000000357 79127 79324 + 155 M 14 I 1 M 184
popper[zn1]1582:

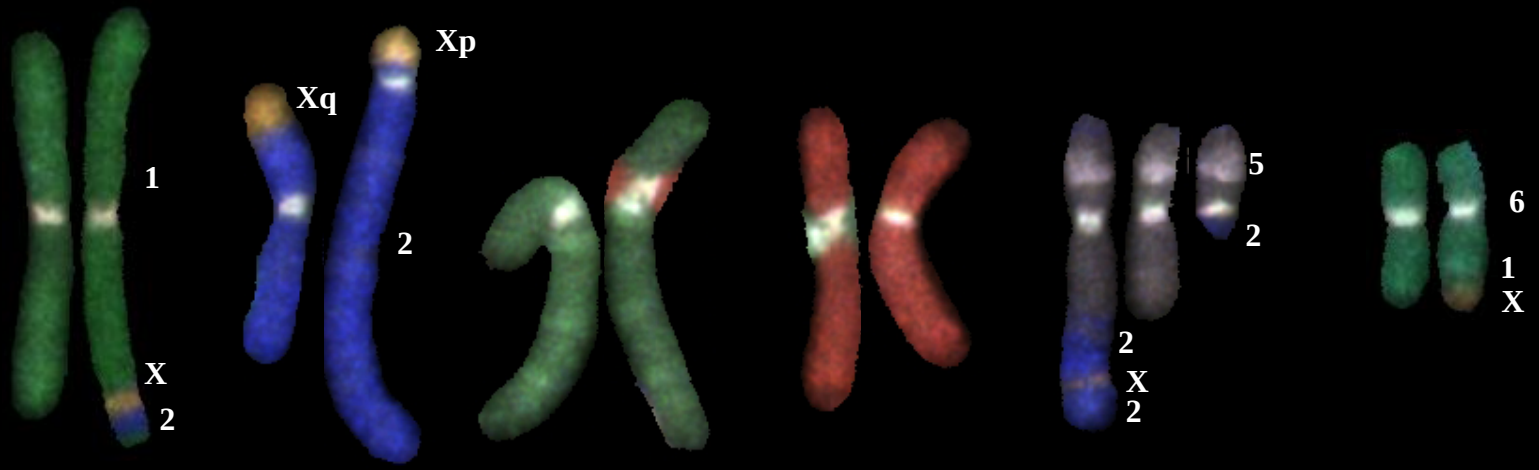
DFTD1


K
I
F1
J
M3
M1
G/H
F
E
F
D
F2
A
M2?
1
der1
der2
3
4
5
der5
6
der6
M4


1
X
2
X
2
5
6
5
X
6
5
2
X?
2

DFTD2


K3
L
J
M
K1/K2
I
J
H
D
F
G
B
M3
M1
M2
der6
der1
der5
1
2
3
4
5
6


1
X
2
Xp
Xq
2
5
2
2
X
2
6
1
X

Acknowledgements:

- Elizabeth Murchuson
- David McBride
- Yong Gu
- Fengtang Yang
- Bronwen Aken
- Mike Stratton

- Ole Schulz-Trieglaff
- Dirk Evers
- David Bentley