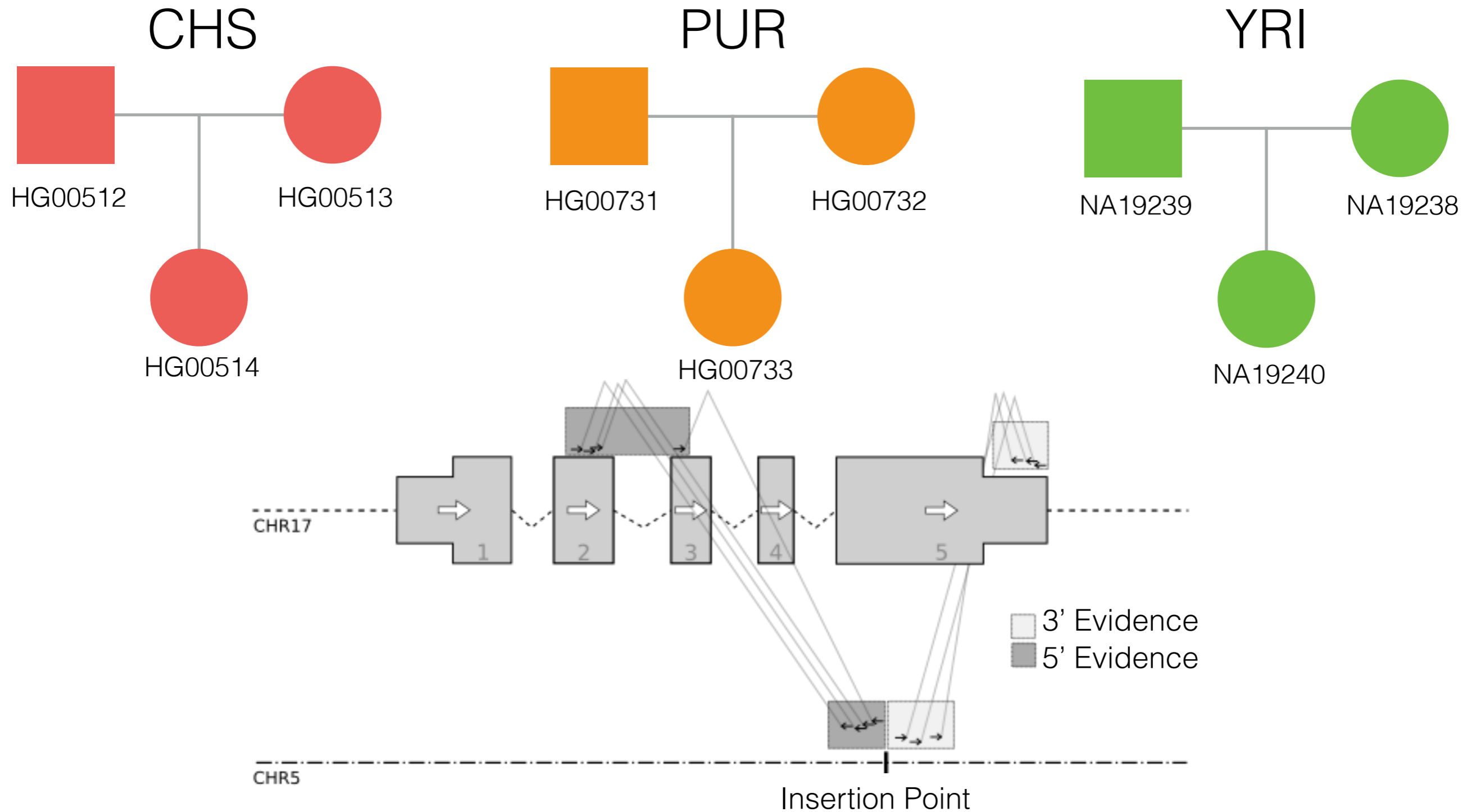# retroCNV (retrodup) calling pipeline
1000g-sv-Trio

"Available data"

- **Low coverage Illumina (phase3)**

- **High Coverage Illumina**

- **PacBio**

- Bionano 90x
  No sequence info
- Jumping library
  Very short reads

- Strand-Seq
  CHS trio
- 10x Genomics
  Unaccessible
- Moleculo
  Not uploaded
- ACGH (array)

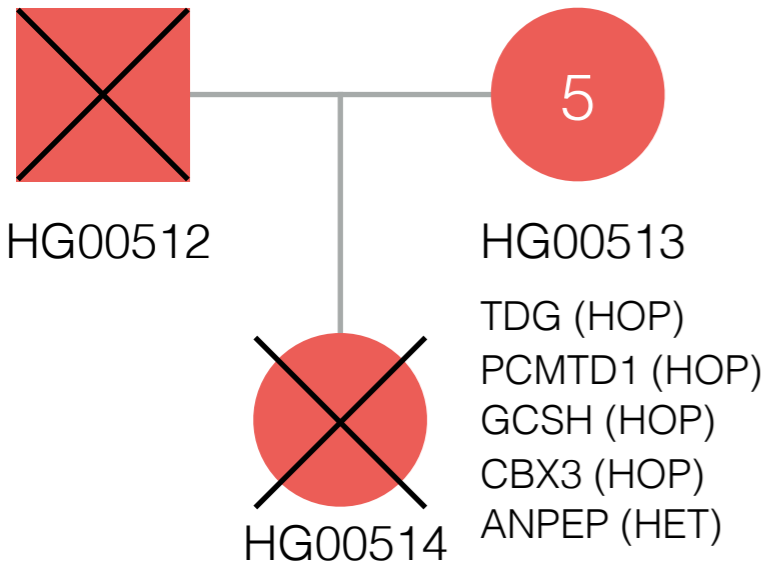- RNA-Seq (PacBio)

- RNA-Seq (Illumina)
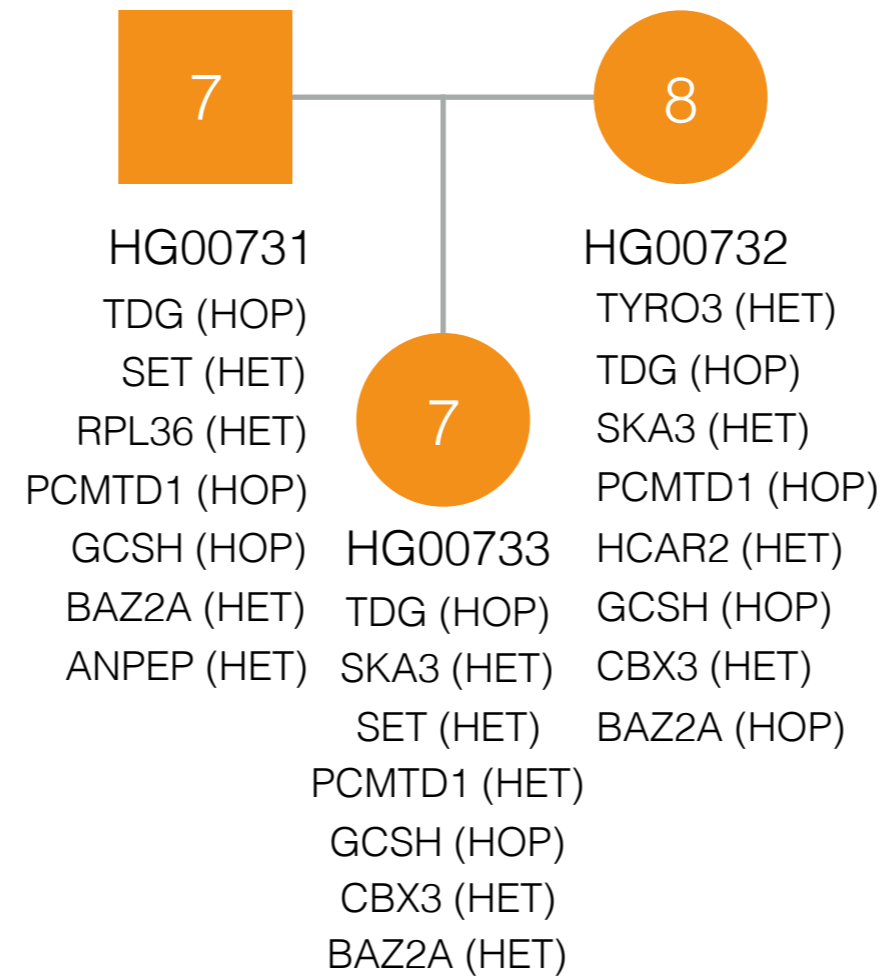
Illumina paired-end alignments

CHS

HG00512   HG00513
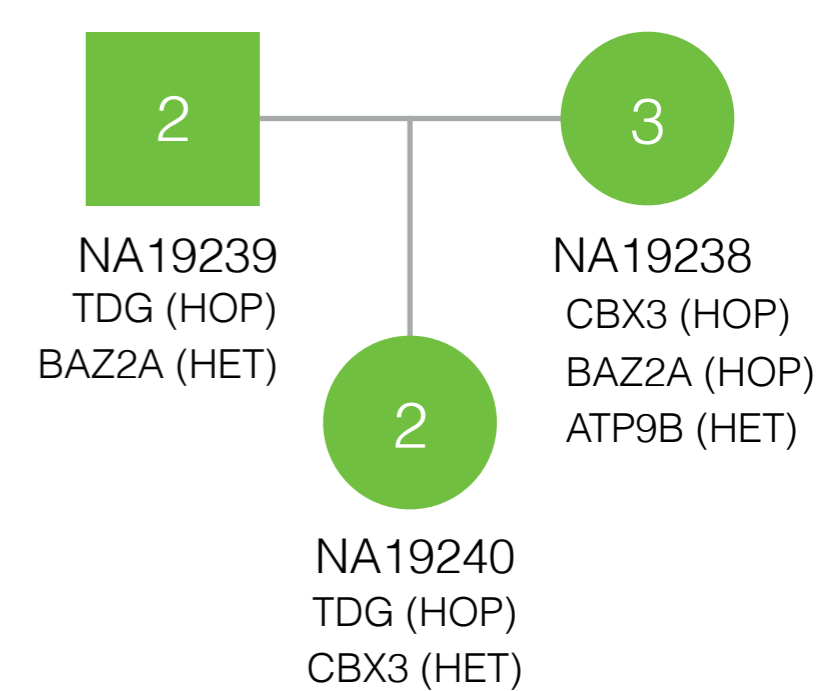
HG00514

PUR

HG00731   HG00732

HG00733

YRI

NA19239   NA19238

NA19240

CHR17

1   2   3   4   5

3' Evidence
5' Evidence

CHR5

Insertion Point

3

1000G Low coverage

(phase3)

CHS

HG00512

HG00514

HG00513

TDG (HOP)
PCMTD1 (HOP)
GCSH (HOP)
CBX3 (HOP)
ANPEP (HET)

PUR

5

7

8

7

HG00731

TDG (HOP)
SET (HET)
RPL36 (HET)
PCMTD1 (HOP)
GCSH (HOP)
BAZ2A (HET)
ANPEP (HET)

HG00733

TDG (HOP)
SKA3 (HET)
SET (HET)
PCMTD1 (HET)
GCSH (HOP)
CBX3 (HET)
BAZ2A (HET)

HG00732

TYRO3 (HET)
TDG (HOP)
SKA3 (HET)
PCMTD1 (HOP)
HCAR2 (HET)
GCSH (HOP)
CBX3 (HET)
BAZ2A (HOP)

YRI

2

3

2

NA19239

TDG (HOP)
BAZ2A (HET)

NA19240

TDG (HOP)
CBX3 (HET)

NA19238

CBX3 (HOP)
BAZ2A (HOP)
ATP9B (HET)

4

High coverage PCR free

CHS

HG00512 — 9

HG00513 — 6

HG00514 — 9

HG00513
ANPEP(HET)
CBX3(HOP)
GCSH(HOP)
PCMTD1(HOP)
TDG(HET)
BAZ2A(HET)

HG00512
ANPEP(HET)
BAZ2A(HET)
CBX3(HOP)
GCSH(HOP)
CMTD1(HOP)
SET(HET)
SET(HOP)
SKA3(HOP)
TDG(HOP)

HG00514
ANPEP(HET)
BAZ2A(HET)
CBX3(HOP)
GCSH(HOP)
PCMTD1(HOP)
SET(HET)
SET(HET)
SKA3(HET)
TDG(HET)

PUR

HG00731 — 8

HG00732 — 10

HG00733 — 6

HG00731
ANPEP(HET)
BAZ2A(HET)
GCSH(HOP)
PCMTD1(HOP)
RPL36(HET)
SET(HET)
TDG(HOP)
SET(HET)

HG00732
BAZ2A(HET)
CBX3(HET)
GCSH(HOP)
HCAR2(HET)
PCMTD1(HOP)
SKA3(HET)
TDG(HOP)
TYRO3(HET)
SET(HET)
SET(HET)

HG00733
BAZ2A(HET)
CBX3(HET)
GCSH(HOP)
PCMTD1(HOP)
SKA3(HET)
TDG(HOP)

YRI

NA19239 — 11

NA19238 — 11

NA19240 — 8

NA19239
TDG(HOP)
BAZ2A(HET)
ATP9B(HET)
CBX3(HET)
GCSH(HOP)
MFF(HET)
PCMTD1(HOP)
RPLP0(HET)
SET(HET)
SET(HET)
TMEM5(HET)

NA19238
ATP9B(HET)
BAZ2A(HET)
CBX3(HET)
GCSH(HOP)
MFF(HET)
PCMTD1(HOP)
RPLP0(HET)
SET(HET)
SET(HET)
TDG(HOP)
TMEM5(HET)

NA19240
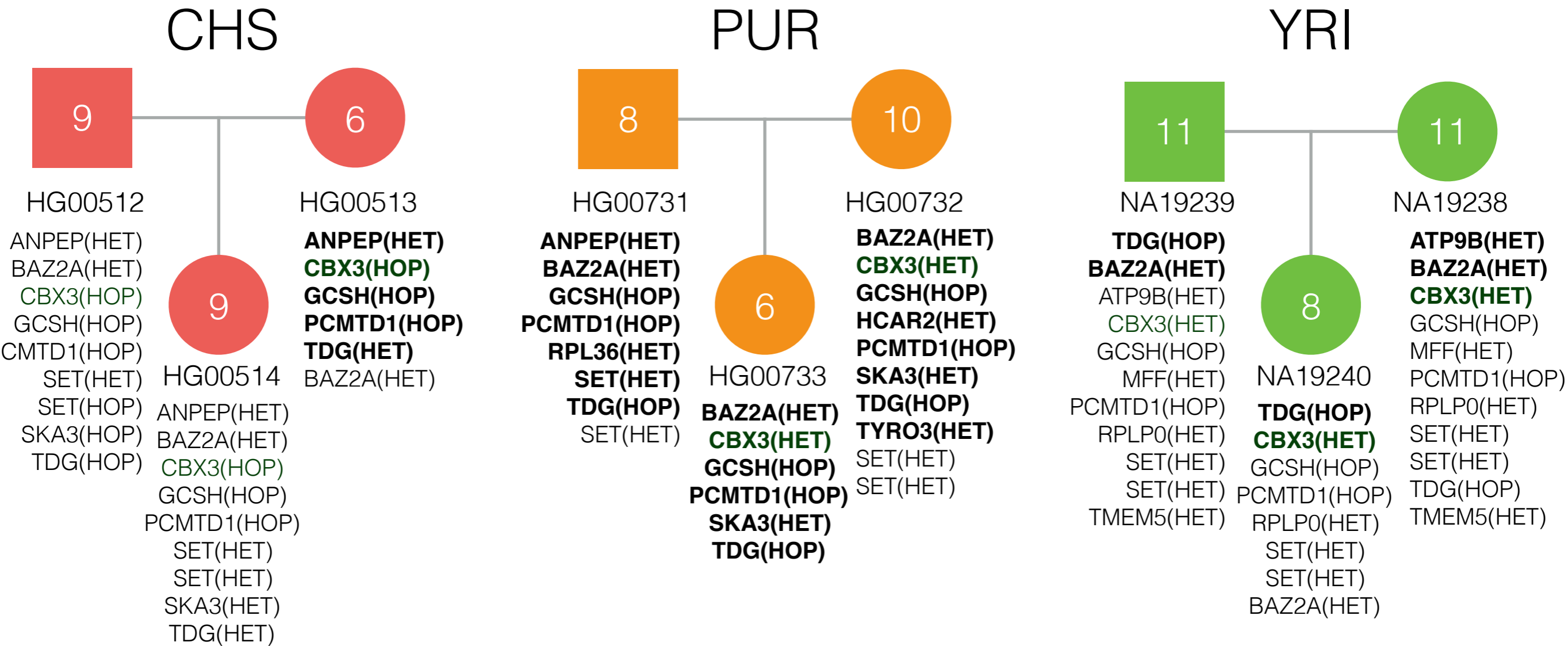**TDG(HOP)**
CBX3(HET)
GCSH(HOP)
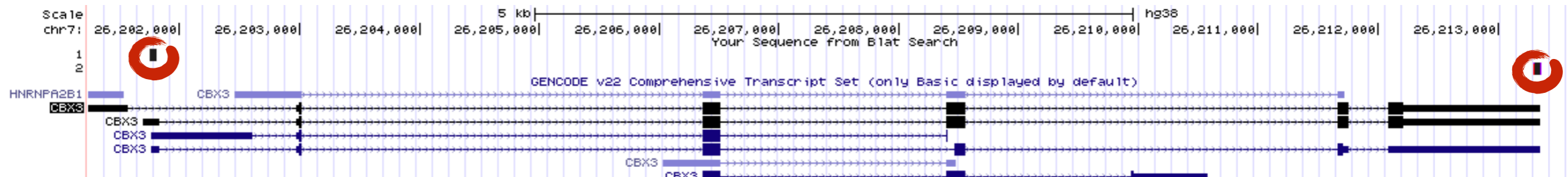PCMTD1(HOP)
RPLP0(HET)
SET(HET)
SET(HET)
BAZ2A(HET)

Breakpoint analysis
Illumina High Coverage

CBX3 inserted into chr15

chr15 40561981 0 313 CTTCGGATGTGGCTTGAGCTGTAGGCGCGGAGGGCCGGAGACGCTGCAGACCCGCGACCCGGAG
chr15 40561992 1 46 ATTTTTTTTTTTAAAGAAATATATAACTATTTATTAACCACTGTTCAGTATTTACAATAAAGTAAAC

TYRO3 inserted into chr13

chr13 43495691 0 28 GCCCCGCCCTCCTCCCTCCTCGCTCGCGGGCCGGGCCCGGCATGGTGCGGCGTCGCCGCC
chr13 43495703 0 12 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
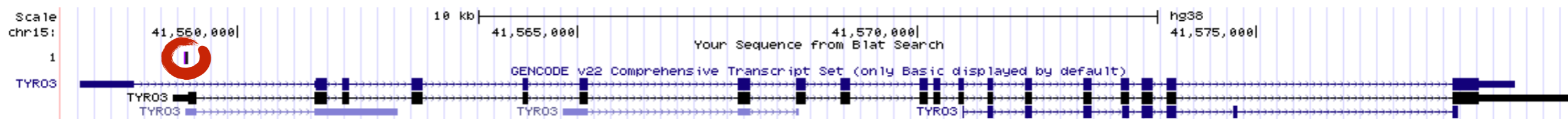
Breakpoints analysis is able to identify all insertions extremities, including Target Site Duplications and poly(A)s

PacBio pipeline?

Using CIGAR code to detect breakpoints with Illumina is easy:

223230194   83   chr15   40561473   60   **18S108M** =   40561044   -537   AAAAAACA

PacBio reads in other hand…

26900   16   chr15   40554595   254
9088S18M1D21M1D23M1D7M1D4M1I10M1D48M1I10M1I3M1D6M1D1M1D26M2I13M1D10M1I2M1D22
M1I3M1D1M1I4M2I18M1I9M2I6M1I7M1I5M1D24M2I7M1I3M1I10M1D3M1I19M1D1M1D13M1D17M1D4M
1D3M1D1M1I9M1I5M1I1M1I23M1I24M1I5M1D5M1I3M1I5M1D10M1I14M1D22M1I2M1D23M1D9M1I4M1I
1M1I7M1D4M1D18M1D5M1D5M1D4M1I1M1D3M1I3M1I8M1D6M1D1M1I5M1D2M1I7M1I27M1D16M1I3M
1I15M1I20M1I6M1I1M1I8M1I2M1D9M1I2M1I5M2D21M1D18M1I10M1I10M1I12M1D9M1D2M1I10M1I28M
1I6M1I9M1I15M1D14M1D4M1D8M1D23M1D17M1D23M1D17M1D6M1D33M1I7M1D14M1I3M2I15M1I3M
1D42M2I1M1I22M1D2M1D16M1I27M1I6M1D2M1D19M1D15M1D2M1D11M2D2M1D15M1I3M1I16M1D7
M1D18M1I1M1I1M2I1M1I1M1I3M1I1M2I1M1I11M1I11M1I2M1I1M1I36M1D7M1D2M1I3M1D6M2I42M1D3
M1I25M1I1M1D2M1D14M1I21M1I6M1D2M1I5M2I3M1D13M1I13M1I4M1I5M1I5M1D10M1I11M1D12M1I6
M1D8M2D4M1I5M1D1M1D10M1D3M1D5M1D2M1I3M1D14M1D2M1D12M1I3M1D3M1D10M1D4M1I15M
1I20M1D10M1D15M1D16M1D10M1D3M1I21M1I13M1D3M1D1M1D6M1I13M2D8M1I1M1I17M1D9M1D1
M1D10M1D15M2D10M1I9M1D3M1D4M1I28M1I14M1I5M1I18M1I7M1D18M4…

4I6M1D14M1D9M1I13M1I4M1I3M1I4M1D17M1D9M1D5M1I10M1I1M1I16M1D8M1D4M1I19M1I13M1I7M1D11M1D6M1D27M1D4M1D20M2I4M2I18M1I2M
1I5M1D17M4I8M1I8M2D3M1I3M1I4M1D23M1D4M1D10M1I17M1D4M1D1M1I4M1D7M1I1M1I1M1I1M1I12M1D17M1I24M2I6M1I14M1D15M1D11M1D4M2
I12M1I35M1I9M1D6M1I18M1I4M1D6M1I21M1D19M2D11M1D38M1D24M1D26M1I14M2I6M1I12M1I40M1D4M1D2M1D3M1D3M1I23M1I1M1D15M1D24M
1I6M1D2M2D7M1D13M1I2M1I34M1I32M2I5M1I4M1I3M1I13M1I27M1I6M1I1M1I1M1I24M1I6M1D23M1D4M1D3M1D1M1D5M1D12M1D37M1I20M1D2M1I
5M1I16M1I9M1I33M1D30M2D20M1D3M1D22M1D7M1D2M1D5M1D6M1D9M1I14M1I20M1I1M1I12M1I11M1I5M1I2M1I3M1D1M1D13M1I3M1I4M1I9M1I3
M1I8M1D4M1D8M1I23M1I22M1D4M1I4M1D7M1D18M1D8M3D4M2D5M1D2M1D7M1I5M1I7M1I7M1D17M1D2M1D12M1D4M1I8M1I7M1I1M1I1M1I9M1
D2M1I5M1D9M1I19M1D13M1D13M1D3M1I5M1I3M1I6M1I1M1I16M1I15M1I23M1I17M1I9M1I1M1I1M1D15M1D2M1I6M1I1M1I10M1I7M1I3M1D26M1I7M1
I21M1D12M1D1M1I2M1I15M1I37M1I3M1D3M1I8M1I37M1I1M2I1M1I12M1I1M1D1M1D12M1I3M1I2M1D12M1D6M1I4M1I7M1D12M1I13M3I3M1D8M1I22
M1D9M1I5M1I10M2D18M1D21M1I4M1D9M1I4M1I4M1D3M1D7M1I2M1I6M1D27M2D7M1D7M1D1M1D3M1D12M1D12M1I7M1I2M1I4M1D2M1D7M1D7M
1D2M1D5M1I12M1D20M1D15M2D5M1D2M1D3M1I12M1D2M1D5M1I3M1I22M1I14M1I1M2I2M1I16M1D9M1I18M1I7M1D10M1D3M1I17M1I1M1I2M1I3M2
I6M1I1M1I1M3I13M1I14M1I15M1D6M1I8M1I1M1I2M1I7M1I24M1D4M1D9M1I8M1I5M1D4M1I2M1D20M1I1M1I25M1D29M1I3M1D6M1I3M1I59M1I3M1D8
M1I1M2I9M1I1M1I1M1D25M1I9M1I7M1I6M1I6M1D5M1I18M1D32M1D2M1D11M1D11M1I46M1I15M1I2M1I3M1I17M1I8M1D6M1D23M1D4M1D2M1D10M
1I11M1I7M1D5M1I6M2I5M3I5M1I3M1D5M1I4M1D2M1I17M1I6M2I1M1I26M2I1M3I4M1D18M1I17M1D11M1I10M1D20M1I6M1I6M2I4M1D3M1I14M1D6M1
D6M1D13M1I16M1I11M1I6M1D39M1I4M1I12M1D1M1I18M1I3M1I5M1D8M1I11M1D4M1D9M1D7M1I5M1I2M1I5M1I16M1I22M1I12M1I4M1I30M1I7M1D1
M1I3M1D14M1I2M1I25M1D7M1D15M1I33M1I4M1I3M1I15M2I15M1I2M1I7M1D7M1I15M1D13M1I5M1D17M2D9M1D1M1D8M1I6M1I5M3I12M1I21M1I11M
1I1M1I3M1I2M1I3M1I3M1D42M1D6M1D1M1D3M1D11M1I5M1I3M1I4M1D2M1D3M1D12M1I18M1D2M2I18M1D5M1D7M1D19M1D12M1D5M1I3M1D25M
1D6M1I8M1I5M1I1M1D6M1D2M1I4M1D9M1I2M1I5M1I23M1D16M1I1M1I26M1D1M1I6M1I7M1D14M1D5M1I16M1D8M1I15M1I15M1I8M1D4M1I2M1I2M1
D3M1D8M9I2M1I13M1D6M1D8M1I15M1I6M1D21M1I6M1I16M1D3M1D2M2I10M2D6M1I18M1I3M1D1M1D9M1D15M1D15M1D4M1D16M1I6M1I28M1I14
M1D17M1D4M1I1M1D8M1D9M1D2M1I3M1I9M1D19M2D13M1I12M1D2M1I4M1I6M1I19M1I5M1D16M1I27M1I5M1I18M1I4M1I13M1D5M1I7M1D36M1D13
M2D20M1I14M1I6M1I20M2D8M1I13M1D2M1D3M1I1M1D7M1D9M1D12M1D7M1D15M1D7M5I2M2I1M5I1M5I1M16I5M4I1M1I2M6I1M4I3M15I2M10I1M8I
1M3I1M3I1M1I2M13I1M1I1M2I2M11I1M1I1M19I2M21I1M1I1M2I1M3I1M1I3M3I1M**1602I**1M2I1M5I1M2I1M1I1M1I13M1I5M1D4M1I7M1I7M1D10M1I5M2I16
M1D24M1I10M1D19M1D5M1I7M2D2M1D3M1D5M1I10M1I9M1I1M1I15M1I9M1D5M2D17M1D7M1I12M1I9M1I6M1D5M1I3M1I6M1D3M1D7M1I6M1I5M1I2
M1D7M1I6M1D12M1D13M1D5M1D1M1D2M1D5M1I5M1D10M1D1M1D1M3D9M1D3M1D1M1D2M1I1M1I10M1D16M1D17M1D9M1I8M1D9M1I7M1D10M1
D6M1D10M1I7M1D14M1D21M1I10M1I12M1I12M1D8M1D16M1D5M1D34M1I4M1D8M1D2M1D3M1I6M1D2M1I3M1D6M1I9M1D14M1I8M1D10M1I8M1I5
M1D30M1D4M1I2M1I14M1I6M1I15M1I16M1I7M1D9M1D23M1D29M1I10M1D9M1I3M1I11M1D26M1I14M1D4M1I6M1D24M1D4M1D16M1I1M1I3M1I6M1I
2M1I2M1I1M1I9M2I8M1I5M1I5M2D5M1I13M1D1M1D6M1I4M1I6M1D14M1I9M1D8M1D1M1D11M1D3M1D1M1D12M1I6M1D10M1I14M1D10M1I20M1D2
M1D2M1I5M1D3M1I4M1I7M1D7M1D13M1I21M1I10M1D6M1I1M1I5M1I21M1D9M1I8M1D9M1I27M1D14M1D33M1D2M1D11M1I16M1I3M1D11M2I4M2I1
M2I9M1I4M1D11M1D4M1D4M1I23M1I1M1D13M1I5M1D2M1D2M1D5M1D11M2D18M1I11M1D9M1I22M1D15M1I1M2D1M1I11M1D4M1D16M1I8M1D2M1
D10M1I11M2I2M1I12M2D8M1I1M1I6M1D10M1D6M1I3M1I10M1I13M1I1M1I16M1D6M1I10M1I8M1D2M1D5M1I8M1D10M1D10M1I4M1D1M1I14M1D10M
1I24M2I9M1D1M1I1M2I9M1D18M1I18M1D5M1I10M1D1M1I2M1I2M1D3M1D10M1I1M1I17M1I3M1D13M1I3M1D7M1I3M3I3M1I11M1D1M1D19M1I18M1D
1M1D20M2I11M1D1M1D4M1I4M2I2M1D26M1I9M1D7M1D1M1D17M1I14M1I5M1D14M1I23M1I18M1D2M1D37M1I7M1D9M1D7M1D1M1I7M1I7M1I16M1I
18M1I4M1D16M1D4M1D37M1I11M1I15M1I4M1D4M1D2M1D2M1D12M1I6M2I15M1I6M1D16M1I5M1I11M1I11M1D1M1D14M1D10M1I14M1I6M1I22M3I...
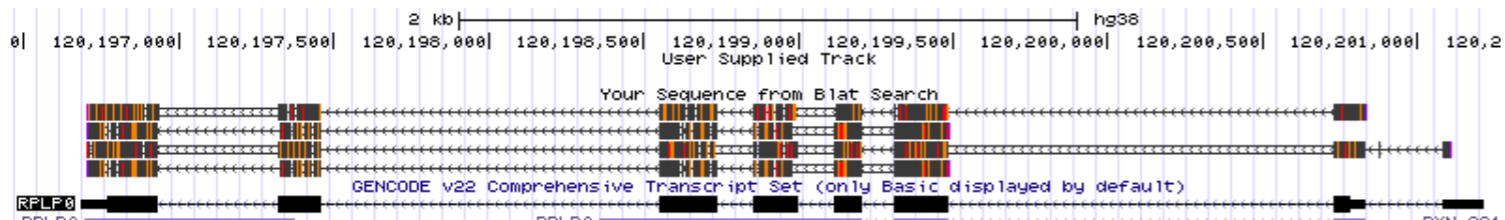
RPLP0

chr11:60274156-60274179

Insertion Point

Scale
chr11:
User Track
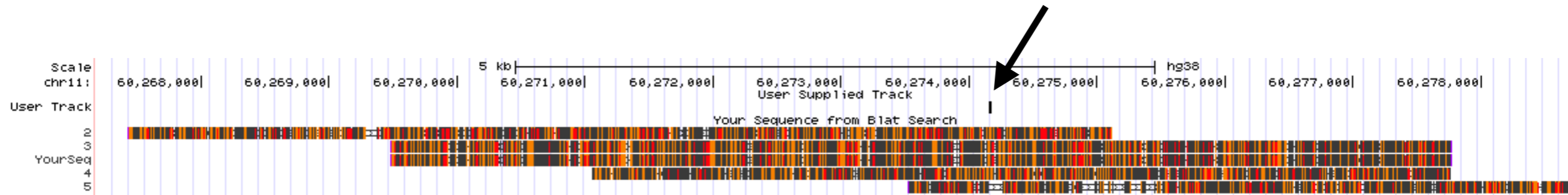
Parental Gene

CTCTCGCCA GGggCgTCCT CaGTGGAAGT GACATCGcTC TTTAAACCCT   5050
GCGTGCAATC CCTGaACGCA CCGCCGTGAT GCCCAGGGAA GACAaGGGCG   5100
ACCTGGAAGT CCAACTACTT CTTAAGaATC ATCCAACTAa TGGATGATTA   5150
TCCgAAATGt TTTCATtGTG GGAGCAGACA ATGTGGGCTC CAAGCAGATG   5200
gCAGCAGATC CGaATGTCCC tTTCGCGGGA AGGctGTGGT GCTGATGGGC   5250
AAGAACACCA TGATGCGCAA GGCCATCCGA GGCACCTGGA AAACAACCCA   5300
GCTCTGagaA AACTGCTGCC TCgttTATCc GGGGGAATGT GGGCTTTGTG   5350
TTCACCAAGG AGGACCgtTC ACTGAaGATC AGGacCATGT TGCctggcCC   5400
...
ATgcagaCCC ATTCTAaTCA TCAACaGGGT AgCAAACGAG TCCTCCTTGT   5850
CTGTGGAGAC GGATTACACC TTCCAaTTGC TGAAgGTCAA GGCCTTCTTG   5900
GCTaCCATCT gGCCTTTGTG GCTGCcTGCC CCTGTGGCTG CTGCACCACA   5950
GCTCTCCaTG CTGCTGCaTG CAGcCCCCAG CTAAGGTTGA AGCCAAGGAA   6000
GAGTaGGAGG AGTCGGcacg aAGGATATGG GATTTGTCTC TTGACTAAat   6050
ACCAAAgcaa AcCCAACTct agGCCAGgtT TTATTTGCAA ACAAGAAATA   6100
AAGGCTTACT TCTTTAAAAA Aaaaaaaaaa aaaaaacttc cgaaactgaa   6150
gagcaaaagg gaaaaaaatg gaaaaaaaga agcagaacaa cccaaagtgc   6200

9