Supplementary Information

Contents **PAGE**

1. Point Mutation Multiplicity Estimation

We have the following likelihood:

$$\Pr(r|M,m,n_s,n_w) \propto \left(\frac{(1-\pi)r}{(1-\pi)(M+m)+2\pi}\right)^{n_s}\left(1-\frac{(1-\pi)r}{(1-\pi)(M+m)+2\pi}\right)^{n_w}, r \in \{1,2,\dots,M\}$$

In this expression we have normal contamination proportion π, major and minor copy numbers M, m, multiplicity r, along with somatically mutated and wild type read counts n_s, n_w.

If we randomly pick a segment from the region concerned, then because normal and cancer segments have copy numbers 2 and $M+m$, and the cells occur in proportions π and $1-\pi$, the relative proportions of normal and cancer segments are 2π and $(M+m)(1-\pi)$.

If the multiplicity is r, then a fraction $r/(M+m)$ of cancer segments have the mutations. The proportions of normal, cancer mutated and cancer wild type segments are 2π, $r(1-\pi)$ and $(M+m-r)(1-\pi)$. The fraction of segments with the somatic mutation is thus $\frac{(1-\pi)r}{(2\pi+(M+m)(1-\pi))}$. We are sampling mutated and un-mutated reads binomially (conditional on the total number of reads), giving the binomial distribution indicated.

2. Solving Edge Consistency Conditions

Formally we describe the edge conservation principle as follows. We let t^{+c}_s and t^{-c}_s count the number of segments (s,c) that are capped by a telomere at the right and left ends, respectively. These can be wild type telomeres or formed somatically during processes such as chromosomal arm loss. We let wild type edge counts g^c_s and denote the number of wild type edges connecting the right ends of minor allele m^c_s of segment (s,c) to segment $(s+1,c)$. G^c_s is the analogous count for the major allele M^c_s. We let somatic edge matrix $h^{++}_{s_1s_2}{}^{c_1c_2}$ count the number of connections between the right ends of segments (s_1,c_1) and (s_2,c_2). Similarly, we also have matrices $h^{+-}_{s_1s_2}{}^{c_1c_2}$, $h^{-+}_{s_1s_2}{}^{c_1c_2}$, and $h^{--}_{s_1s_2}{}^{c_1c_2}$ where + and – indicate right and left ends respectively. Note that we have the following matrix symmetries:

$$h^{+-}_{s_1s_2}{}^{c_1c_2} = h^{-+}_{s_2s_1}{}^{c_2c_1},\ h^{++}_{s_1s_2}{}^{c_1c_2} = h^{++}_{s_2s_1}{}^{c_2c_1} \text{ and } h^{--}_{s_1s_2}{}^{c_1c_2} = h^{--}_{s_2s_1}{}^{c_2c_1}$$

Both ends of all segments must attach to its wild type partner, be somatically attached to another segment, or be capped with a telomere. Accounting for all copies of such segment ends provides the following system of equations.

Left (5') Conservation ($s>1$):

$$m^c_s = \alpha^c_{s-1}g^c_{s-1} + (1-\alpha^c_{s-1})G^c_{s-1} + (\alpha^c_{s-1}(1-\beta^c_{s-1}) + (1-\alpha^c_{s-1})\beta^c_{s-1})\left\{\sum_{\{s_1,c_1\}\neq\{s,c\}} h^{--}_{s_1s}{}^{c_1c} + \sum_{\{s_1,c_1\}} h^{+-}_{s_1s}{}^{c_1c} + 2h^{--}_{ss}{}^{cc} + t^{-c}_s\right\},$$

$$M_s^c = (1-\alpha_{s-1}^c)g_{s-1}^c + \alpha_{s-1}^c G_{s-1}^c$$
$$+ ((1-\alpha_{s-1}^c)(1-\beta_{s-1}^c) + \alpha_{s-1}^c\beta_{s-1}^c)\left\{\sum_{\{s_1,c_1\}\neq\{s,c\}} h^{--c_1c}_{\ \ s_1s} + \sum_{\{s_1,c_1\}} h^{+-c_1c}_{\ \ s_1s} + 2h^{--cc}_{\ \ ss} + t^{-c}_{\ s}\right\}.$$

Right (3') Conservation($s < S_c$):

$$m_s^c = g_s^c + (1-\beta_s^c)\left\{\sum_{\{s_2,c_2\}\neq\{s,c\}} h^{++c_2c}_{\ \ s_2s} + \sum_{\{s_2,c_2\}} h^{-+c_2c}_{\ \ s_2s} + 2h^{++cc}_{\ \ ss} + t^{+c}_{\ s}\right\},$$

$$M_s^c = G_s^c + \beta_s^c\left\{\sum_{\{s_2,c_2\}\neq\{s,c\}} h^{++c_2c}_{\ \ s_2s} + \sum_{\{s_2,c_2\}} h^{-+c_2c}_{\ \ s_2s} + 2h^{++cc}_{\ \ ss} + t^{+c}_{\ s}\right\}.$$

Note that elements $h^{++cc}_{\ \ ss}$ and $h^{--cc}_{\ \ ss}$ are counted twice to reflect the fact that we have a single rearrangement between two copies of the same segment and so need to account for two copies of the segment being conserved.

We assume that the wild type telomeres are not involved in rearrangements. Then the number of wild type ends of the chromosomes must satisfy:

$$m_1^c + M_1^c = t^{-c}_{\ 1},$$

$$m_{S_c}^c + M_{S_c}^c = t^{+c}_{\ S_c}.$$

3. Maximum Likelihood Edge Consistency Solution

The edge consistency conditions provide candidate solutions counting the number of connections in the genome. There may be more than one solution to this question. To identify the more probable solution we can use frequency of reads bridging each connection to derive the maximum likelihood solution.

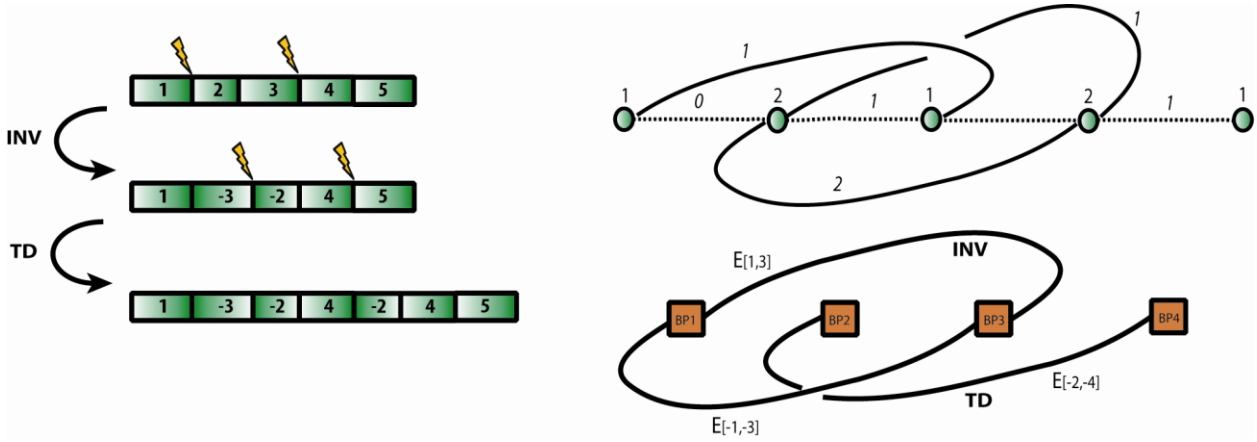
Firstly we can consider the frequencies with which these occur. For every somatic edge count $h^{--c_1c_2}_{\ \ s_1s_2}$, $h^{-+c_1c_2}_{\ \ s_1s_2}$, $h^{+-c_1c_2}_{\ \ s_1s_2}$ and $h^{++c_1c_2}_{\ \ s_1s_2}$ there are counts of reads bridging the associated breakpoint $n^{--c_1c_2}_{\ \ s_1s_2}$, $n^{-+c_1c_2}_{\ \ s_1s_2}$, $n^{+-c_1c_2}_{\ \ s_1s_2}$ and $n^{++c_1c_2}_{\ \ s_1s_2}$ respectively. Wild type edge counts G_s^c, g_s^c are both associated with the same type of wild type read count n_s^c. Each of the somatic counts can be approximated as a Poisson distribution with mean rates $\rho h^{--c_1c_2}_{\ \ s_1s_2}$, $\rho h^{-+c_1c_2}_{\ \ s_1s_2}$, $\rho h^{+-c_1c_2}_{\ \ s_1s_2}$ and $\rho h^{++c_1c_2}_{\ \ s_1s_2}$ where ρ is a parameter representing the mean coverage across a single position per copy number. Wild type counts have mean rate $\rho(g_s^c + G_s^c)$. We then obtain a log-likelihood for the data of:

$$LL = max_\rho\left\{\sum_{s,c}\left[\text{LogPoiss}_{n_s^c}(\rho(g_s^c + G_s^c))\right] + \sum_{s_1,s_2,c_1,c_2}\text{LogPoiss}_{n^{++c_1c_2}_{\ \ s_1s_2}}\left(\rho h^{++c_1c_2}_{\ \ s_1s_2}\right)\right\}$$

The log-likelihoods for each possible solution arising from the edge consistency conditions can then be calculated and the solution with maximum likelihood selected.

4. Transformation Orientations.

When classifying the transformations from the components of the somatic graph we need to consider the different orientations at breakpoints, corresponding to cases when the genome at the breakpoint is inverted prior to the transformations. We highlight this with the following example.



Consider the (algebraic) chromosome [1 2 3 4 5]. We first perform an inversion (INV) using the 1st and 3rd breakpoints BP1 and BP3 to give [1 -3 -2 4 5]. We then perform a tandem duplication (TD) with 2nd and 4th breakpoints BP2 and BP4 to give [1 -3 -2 4 -2 4 5]. The corresponding somatic graph with two components can be seen above, one arising from the TD the other from the INV. Note that when comparing the component corresponding to the TD to that of Figure 2, the edge at the left node points in opposite direction. This is precisely because the left breakpoint was inverted by the INV.

We formally recognize this as follows. When we compare the {BP1, BP3} component to Figure 2, we find a match to INV. For the remaining component we need to consider the different orientations. The connection matrices for this component of the somatic graph have non-zero elements $I^{--}(\mathrm{BP2},\mathrm{BP4}) = I^{--}(\mathrm{BP4},\mathrm{BP2}) = 1$. This does not match anything in Figure 2. Inverting breakpoint BP2 gives $I^{+-}(\mathrm{BP2},\mathrm{BP4}) = I^{-+}(\mathrm{BP4},\mathrm{BP2}) = 1$, which matches the TD connection matrices. Inverting breakpoint BP4 gives $I^{-+}(\mathrm{BP2},\mathrm{BP4}) = I^{+-}(\mathrm{BP4},\mathrm{BP2}) = 1$, which matches the DEL or UT connection matrices. We need more information to decide which one of the three operations is correct. The path connectivity resolves this.

The genomic connection between BP2 and BP4 where the genome has not been flipped at these positions is represented as $2^{+}\sim4^{-}$. If the genome at BP2 is flipped (for the TD match) this becomes $2^{-}\sim4^{-}$, if BP4 is flipped (for the DEL/UT match), $2^{+}\sim4^{+}$. Initially we have [1 2 3 4 5]. We need either the genome at BP2 to be flipped (for a TD) or BP4 to be flipped (for a DEL/UT). Neither is flipped so we implement the INV first to give [1 -3 -2 4 5]. We now see that the left side of BP2 (segment 2) is connected to the left side of BP4 ($2^{-}\sim4^{-}$), this only matches the TD option and we have pinpointed the correct transformation.

5. Ordering Transformations

We give a formal description of the algorithm. Examples follow the description.

I. The Algorithm

We have an order of transformations $T_1 < T_2 < T_3 < \cdots < T_R$ to implement on a set of C diploid chromosomes $C_1^M, C_1^F, C_2^M, C_2^F, \ldots, C_c^F$ in an attempt to recapitulate the evolution of the observed cancer genome. We proceed as follows.

Step 1: Simulate by Component. Each component of the allelic graph consists of a subset of chromosomes connected by a subset of somatic transformations. Thus for each Allelic Graph component $i = 1, \ldots, I$ we have a subset of chromosomes $\{C_{c_1^i}^{p_1^i}, C_{c_2^i}^{p_2^i}, C_{c_3^i}^{p_3^i}, \ldots, C_{c_{i_{max}}^i}^{p_{i_{max}}^i}\}$, where $p_j^i \epsilon \{M, F\}$ denotes the allele for the j[th] chromosome of the i[th] Allelic Graph component, and $c_j^i \epsilon \{M, F\}$ denotes the chromosomal number of the j[th] chromosome of the i[th] Allelic Graph component. We also have a subset of R_i transformations $\{T_{r_1^i} < T_{r_2^i} < \cdots < T_{r_{R_i}^i}\}$.

Step 2: Initialize the chromosomes. If S_j^i denotes the corresponding number of segments we construct algebraic strings to represent each chromosome at the beginning of evolution:

$$A_j^i = [1_{c_j^i}^{p_j^i}, 2_{c_j^i}^{p_j^i}, 3_{c_j^i}^{p_j^i}, \ldots, {S_j^i}_{c_j^i}^{p_j^i}]$$

Step 3: Sequentially Perform Transformations. Each transformation $T_{r_2^i}$ acts on a specific set of breakpoints and modifies the genome as indicated in Figure 1. When there is more than one copy of a breakpoint we try the transformation at all copies of the breakpoint. These have the following string operations for each transformation (we simplify the notation):

A: Terminal Deletion at breakpoint $[k, k+1]$:

$[a, \ldots., k, k+1, \ldots, z]$ becomes $[a, \ldots, k]$

B: Breakage-Fusion-Bridge at breakpoint $[k, k+1]$:

$[a, \ldots, k, k+1, e, f, \ldots, z]$ becomes

$[-z, \ldots, -f, -e, -(k+1), k+1, e, f, \ldots, z]$

C: Interstitial Deletion from $[k, k+1]$ to $[l, l+1]$:

The breakpoints must occur in the same chromosome in the order indicated.

Then $[a, \ldots, k, k+1, \ldots, l, l+1, \ldots, z]$ becomes $[a, \ldots, k, l+1, \ldots, z]$

D: Tandem Duplication from $[k, k+1]$ to $[l, l+1]$:

The breakpoints must occur in the same chromosome in the order indicated.

Then $[a, \dots, k, k+1, \dots, l, l+1, \dots, z]$ becomes

$$[a, \dots, k, k+1, \dots, l, k+1, \dots, l, l+1, \dots, z]$$

E: Inverted Duplication from $[k, k+1]$ to $[l, l+1]$:

The breakpoints must occur in the same chromosome in the order indicated.

Then $[a, \dots, k, k+1, \dots, l, l+1, \dots, z]$ becomes

$$[a, \dots, k, k+1, \dots, l, -l, \dots, -(k+1), l+1, \dots, z]$$

F: Inversion from $[k, k+1]$ to $[l, l+1]$:

The breakpoints must occur in the same chromosome in the order indicated.

Then $[a, \dots, k, k+1, \dots, l, l+1, \dots, z]$ becomes

$$[a, \dots, k, -l, \dots, -(k+1), l+1, \dots, z]$$

G: Translocation between $[k, k+1]$ and $[l, l+1]$:

The breakpoints must occur on distinct chromosomes.

Then $[a, \dots, k, k+1, \dots, z]$ and $[a', \dots, l, l+1, \dots, z']$ becomes

$[a, \dots, k, l+1, \dots, z']$ and $[a', \dots, l, k+1, \dots, z]$

H: Unbalanced Translocation between $[k, k+1]$ and $[l, l+1]$:

The breakpoints must occur on distinct chromosomes.

Then $[a, \dots, k, k+1, \dots, z]$ and $[a', \dots, l, l+1, \dots, z']$ becomes $[a, \dots, k, l+1, \dots, z']$

I: Insertion of the segment from $[k, k+1]$ to $[l, l+1]$ into position $[u, u+1]$.

The breakpoints $[k, k+1]$ and $[l, l+1]$ must occur in the same chromosome

in the order indicated.

Then $[a, \dots, k, k+1, \dots, l, l+1, \dots, z]$ and $[a', \dots, u, u+1, \dots, z']$ becomes

$[a, \dots, k, l+1, \dots, z]$ and $[a', \dots, u, k+1, \dots, l, u+1, \dots, z']$

J: Chromosomal Duplication.

Simply duplicate $[a, \dots, z]$ into two strings

K: Chromosomal Deletion.

Remove string $[a, \dots, z]$.

Step 4: Count Allelic Copy number. We now count the number of copies of each allele of each segment of each chromosome and compare to the observed allelic copy number. If we don't have a match we reject the simulated evolution.

II: HCC1187 Evolution

For sample HCC1187 (Figure 2iii) we have rearrangements between chromosome 1 and 6. Each chromosomal region has three separate copy number segments. The corresponding allelic graph thus has three pairs of nodes for each chromosome. For chromosome 1, the copy numbers of the major alleles are 2, 4 and 2 respectively, and for the minor allele we have 0, 2 and 2 respectively. The segments of chromosome 6 have the same values in reverse order. These are the values associated with the nodes.

The first segment chromosome 1 has allelic copy numbers of 2 and 0, the second 4 and 2. From the principle of allelic copy number conservation the somatic rearrangement can involve only one of each pair of these alleles, and the remaining alleles must have no change in copy number. This means there must be a wild-type edge joining the two segments of equal copy number 2. There is a nominal wild-type edge joining the other two alleles (the ones implicated in somatic rearrangement), although this connection is not actually represented in the genome since the allelic copy number of the first node is 0. The third segment of chromosome 1 has allelic copy numbers 2 and 2. There must be a wild-type edge joining the second segment with copy number 2 to one of these (it is of no consequence which one). Chromosome 6 has the same connections in reverse order.

We have two rearrangements associated with copy number changes (indicated in red), one forming connection $[2_1,2_6]$ and one forming $[2_1,2_6]$. These can be unambiguously assigned to the parental alleles with the major copy number of 4 on each chromosome since these are the alleles associated with changes in copy number. We thus find that $\beta_1^1 = 0, \beta_2^1 = \beta_1^6 = \beta_2^6 = 1$. This results in the two somatic edges between the two chromosomes. Note that without further analysis we see that the graph has three components; the two simple components each represent two wild type copies of chromosomes 1 and 6, and a rearranged component involving the other parental chromosomes. This demonstrates the power and utility of the allelic copy number conservation principle.

The major node of segment 2_1 has allelic copy number 4 associated to it. The wild type edge $[2_1,3_1]$ accounts for both copies of 3_1 and so has two copies associated to it. The somatic edge $[2_1,2_6]$ extending to the right thus accounts for the remaining two copies of 2_1 and so also has two copies. The wild type edge $[1_1,2_1]$ on the other side of 2_1 has no copies, because there are is zero allelic copy number associated to 1_1. The somatic edge $[2_6,2_1]$ extending to the left of 2_1 thus accounts for all four copies of 2_1 and so is associated to number four. This completes the allelic graph.

For the Somatic Graph, we have four segmental regions and so three breakpoints represented by the three nodes. Rearrangement [2,-2] connects the right side of a copy of segment 2 to the same end of an identical segment. This implicates the 2nd breakpoint twice so we have a loop extending in a rightward

direction from this node. Rearrangement [-2,3] connects the left side of a copy of the second segment to the left side of one of the third segments, implicating the 1st and 2nd breakpoints. We then have an edge between the two corresponding nodes, each end extending left from the node. Finally, rearrangement [3,-3] joins the right side of the third segment back to the same region. This implicates the third breakpoint with a loop in a rightward direction. Note that the resulting Somatic Graph of Figure 1Ciii has two components. We see from the evolution depicted in Figure 1A that the component with two nodes and two edges corresponds to the ID, generating two rearrangements and two breakpoints, whereas the component with one node and one edge corresponds to the BFB cycle.

To simulate the evolution we proceed as follows. The Allelic Graph (Figure 2Biii) has three components, so we construct. There are two simple components, one involving chromosome 1 and one involving chromosome 6. The remaining component involves the other parental chromosomes of chromosomes 1 and 6. We thus initialize as:

Component 1: $[1_1^F, 2_1^F, 3_1^F]$

Component 2: $[1_6^F, 2_6^F, 3_6^F]$

Component 3: $[1_1^M, 2_1^M, 3_1^M]$ and $[1_6^M, 2_6^M, 3_6^M]$

The first component has no transformations involving breakpoints. The allelic graph has two telomeric nodes each with copy numbers of 2. We thus have two chromosomes to construct. We thus require a chromosomal duplication (CD). This gives us 2x$[1_1^F, 2_1^F, 3_1^F]$.

The same applies to the second component, giving 2x$[1_6^F, 2_6^F, 3_6^F]$.

The third component involves three transformations. The first involves breakpoints $[1_1^M, 2_1^M]$ and $[2_6^M, 3_6^M]$ and is a UT, a TD or a DEL. The latter two require the breakpoints in the same chromosome, which is not the case (they are in different chromosomal strings), so the only event can be a UT, which gives us chromosome $[1_6^M, 2_6^M, 2_1^M, 3_1^M]$. The second similarly involves a UT, TD or DEL. We are within a chromosome so it is either a TD or a DEL. The second transformation has a single rearrangement represented by connection matrix element $I^{+-}([2_1, 3_1], [1_6, 2_6]) = 1$. A DEL would require a genomic path from leftward breakpoint $[2_1^M, 3_1^M]$ to rightward breakpoint $[1_6^M, 2_6^M]$. The current form $[1_6^M, 2_6^M, 2_1^M, 3_1^M]$ does not allow this. Permuting breakpoints, this matches TD element in Figure 3; $I^{-+}([1_6, 2_6], [2_1, 3_1]) = 1$. This requires a genomic path from leftward breakpoint $[1_6^M, 2_6^M]$ to rightward breakpoint $[2_1^M, 3_1^M]$, giving chromosome $[1_6^M, 2_6^M, 2_1^M, 2_6^M, 2_1^M, 3_1^M]$. The final operation, restoring the chromosomal count, is a chromosomal duplication (CD) giving 2x$[1_6^M, 2_6^M, 2_1^M, 2_6^M, 2_1^M, 3_1^M]$.

When we examine the number of copies of each segment we find that:

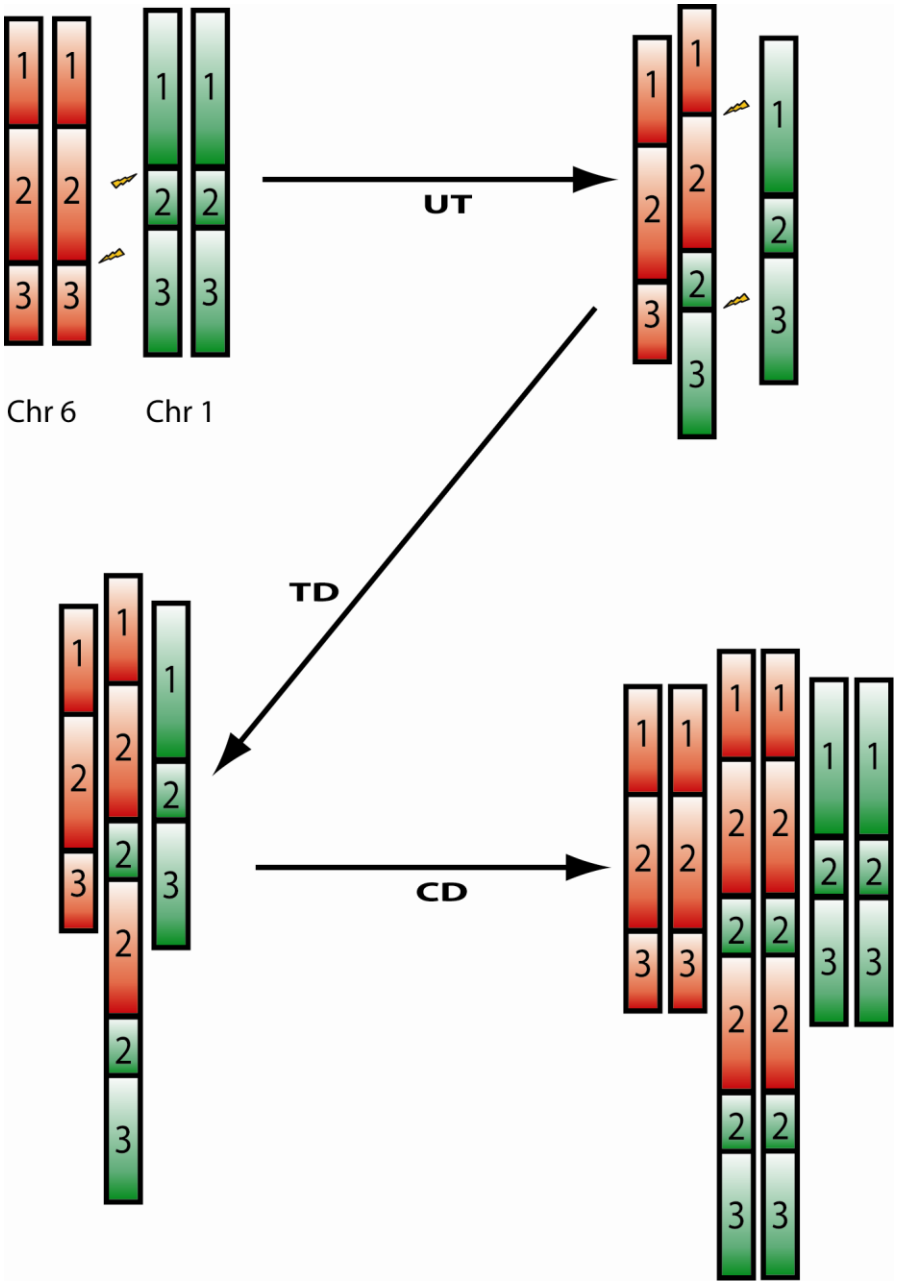
$[1_1^F, 2_1^F, 3_1^F]$ occurs with counts [2,2,2]

$[1_1^F, 2_1^F, 3_1^F]$ occurs with counts [2,2,2]

$[1_1^M, 2_1^M, 3_1^M]$ occurs with counts [0,4,2]

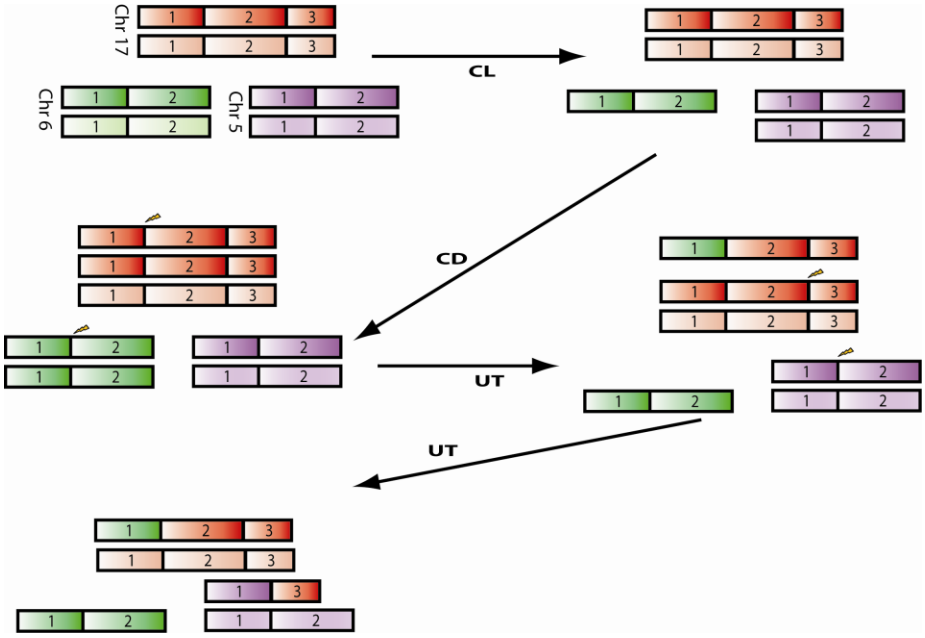
$[1_1^M, 2_1^M, 3_1^M]$ occurs with counts [2,4,0]

Because this agrees with the observed allelic copy number, we conclude we have a valid evolution of events. We also have the 6 chromosomal contigs produced by this evolution. These are indicated pictorially below.

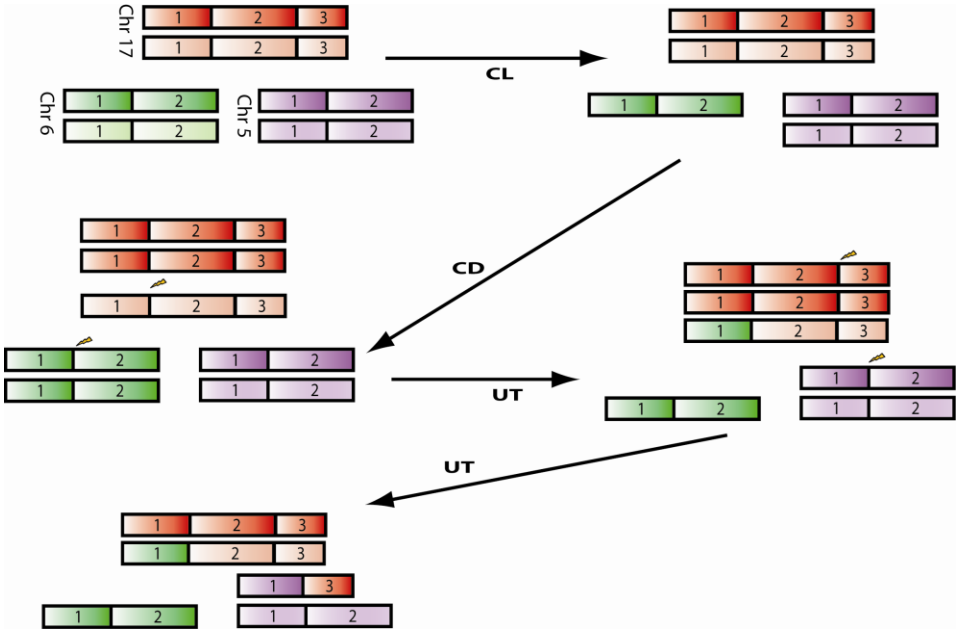


III. Evolution 2: PD3904 1st cluster

For the Allelic Graph of Figure 2Bi we have the following evolution:

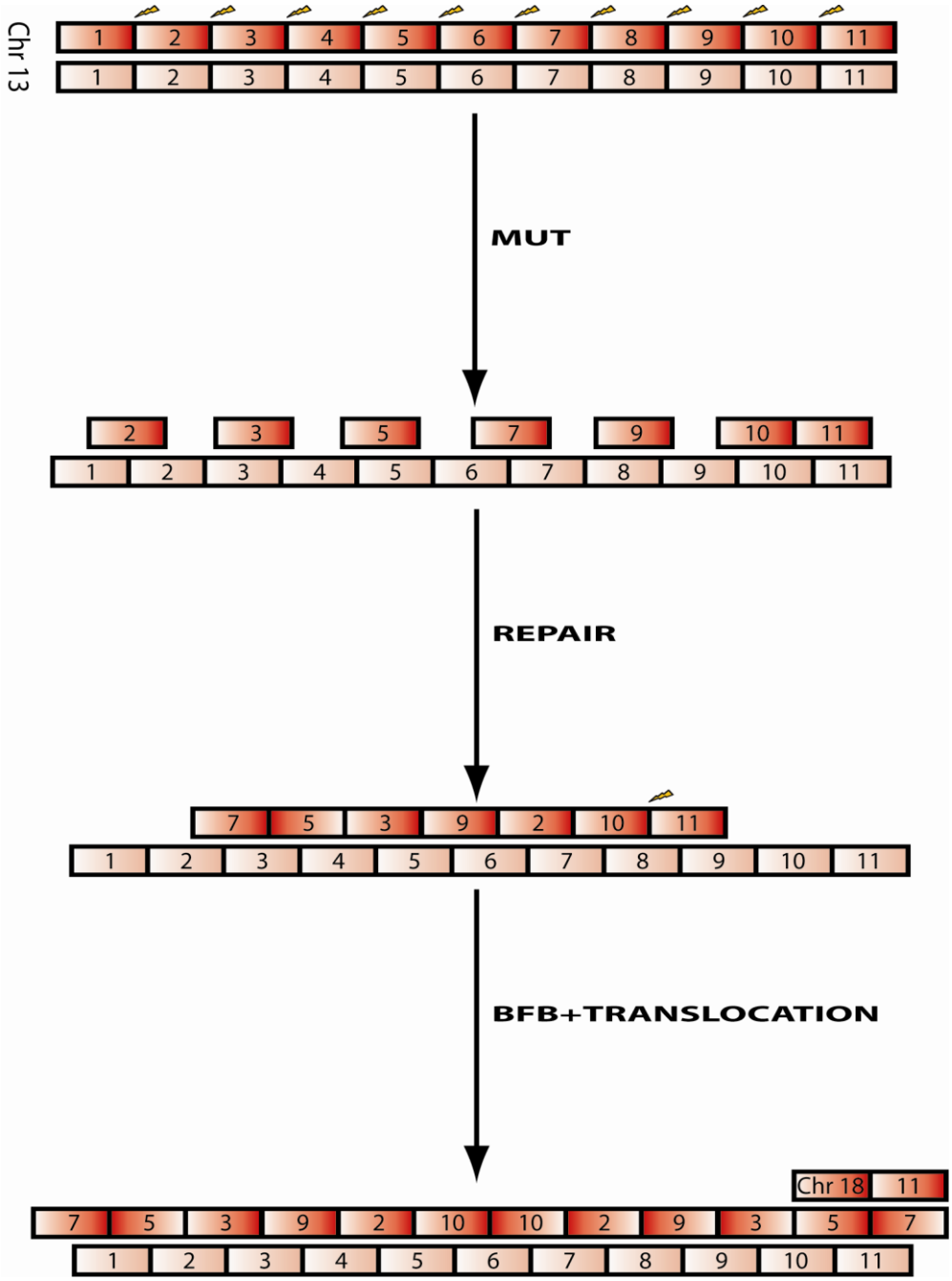


If the alternative Allelic Graph is used (grey lines of Figure 2Bi), we get the slightly different contigs, although the order of events is the same. Both solutions are consistent with the data and we cannot, without further experiment determine the correct configuration.

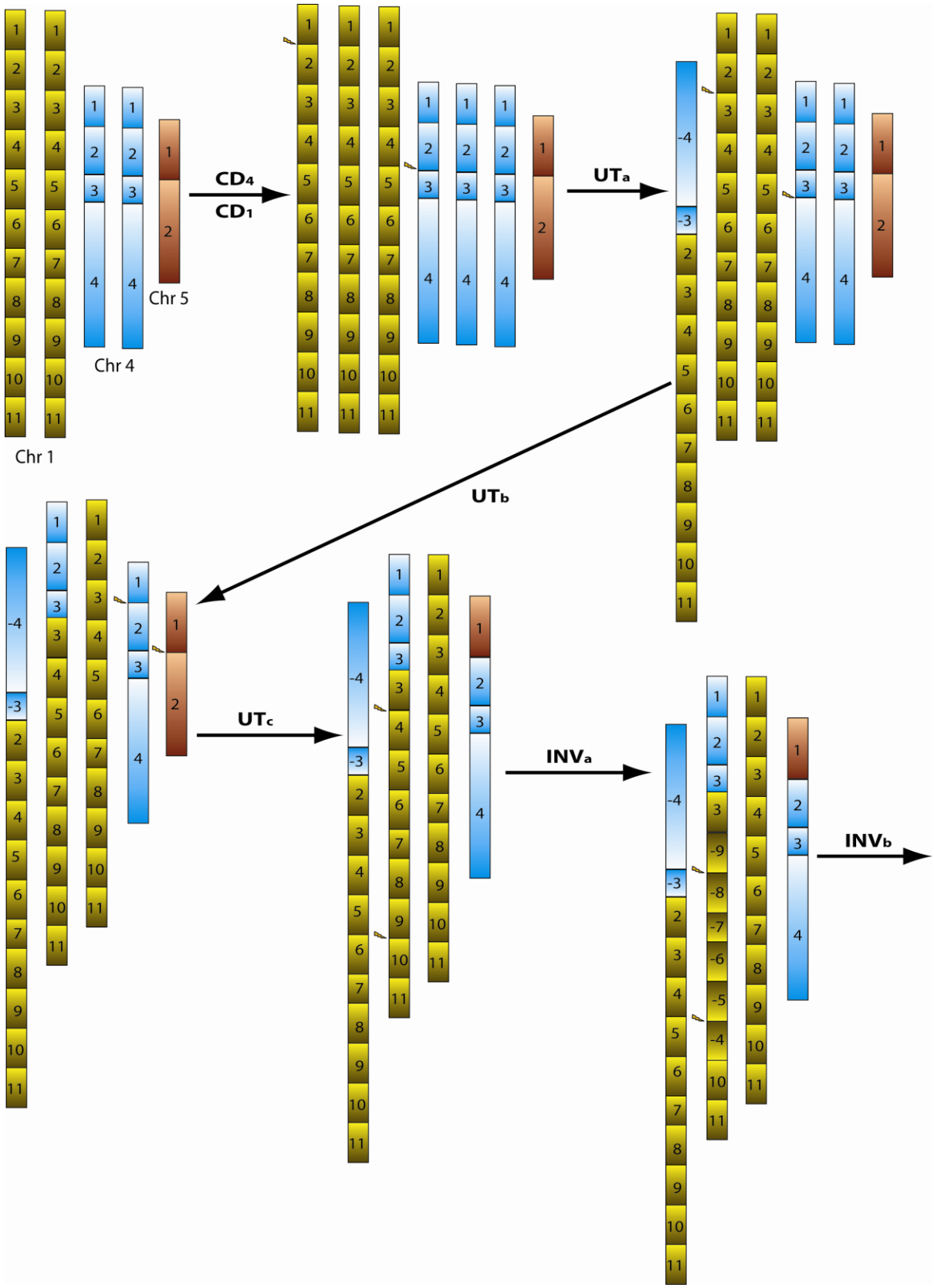


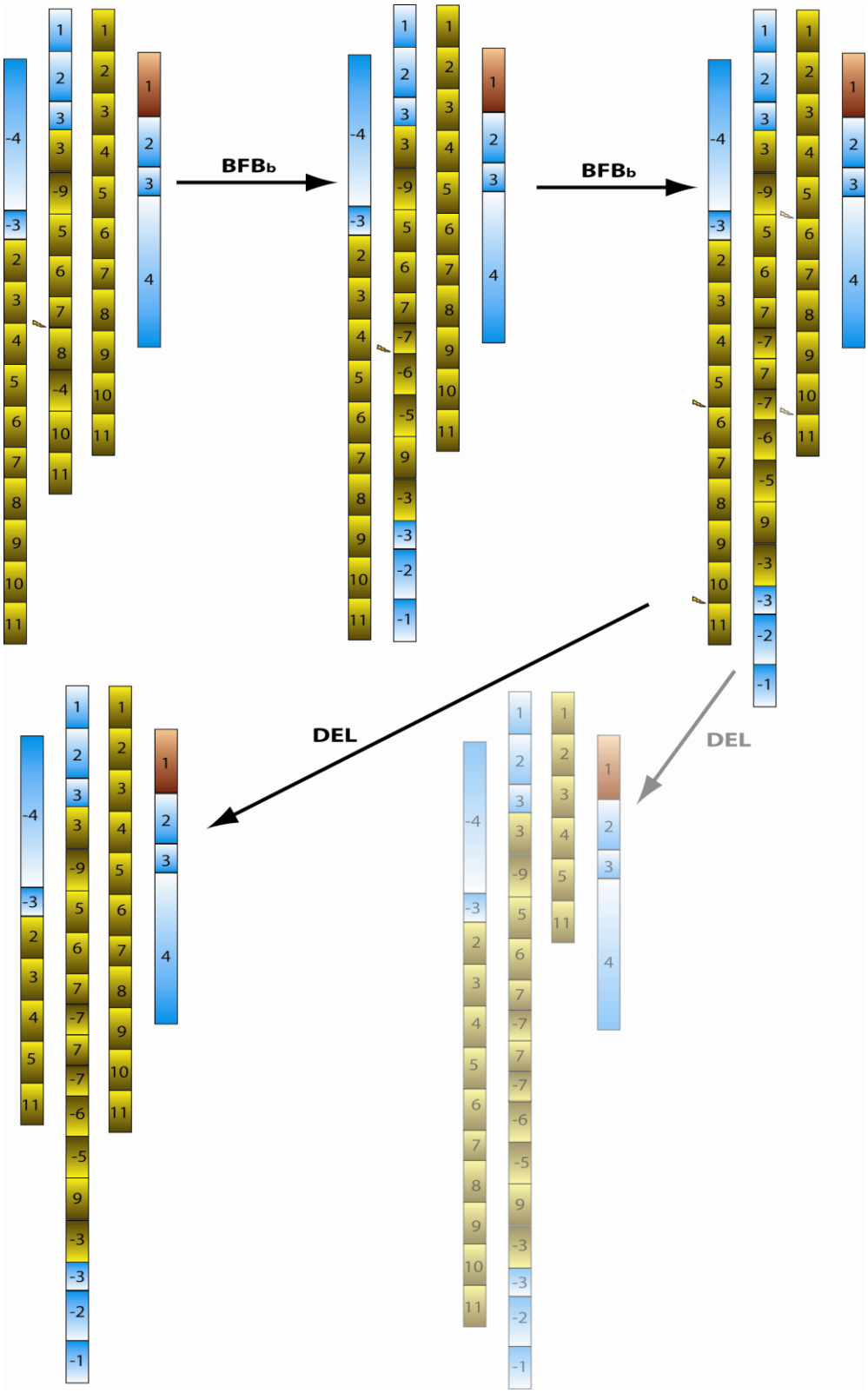
IV. Evolution 3: PD3904 2nd Cluster

For the rearrangement cluster of Figure 2Bii the somatic graph components were not of standard form, where graph walking gave a palindromic contig suggesting a BFB followed a complex rearrangement event. The evolution is then as follows:



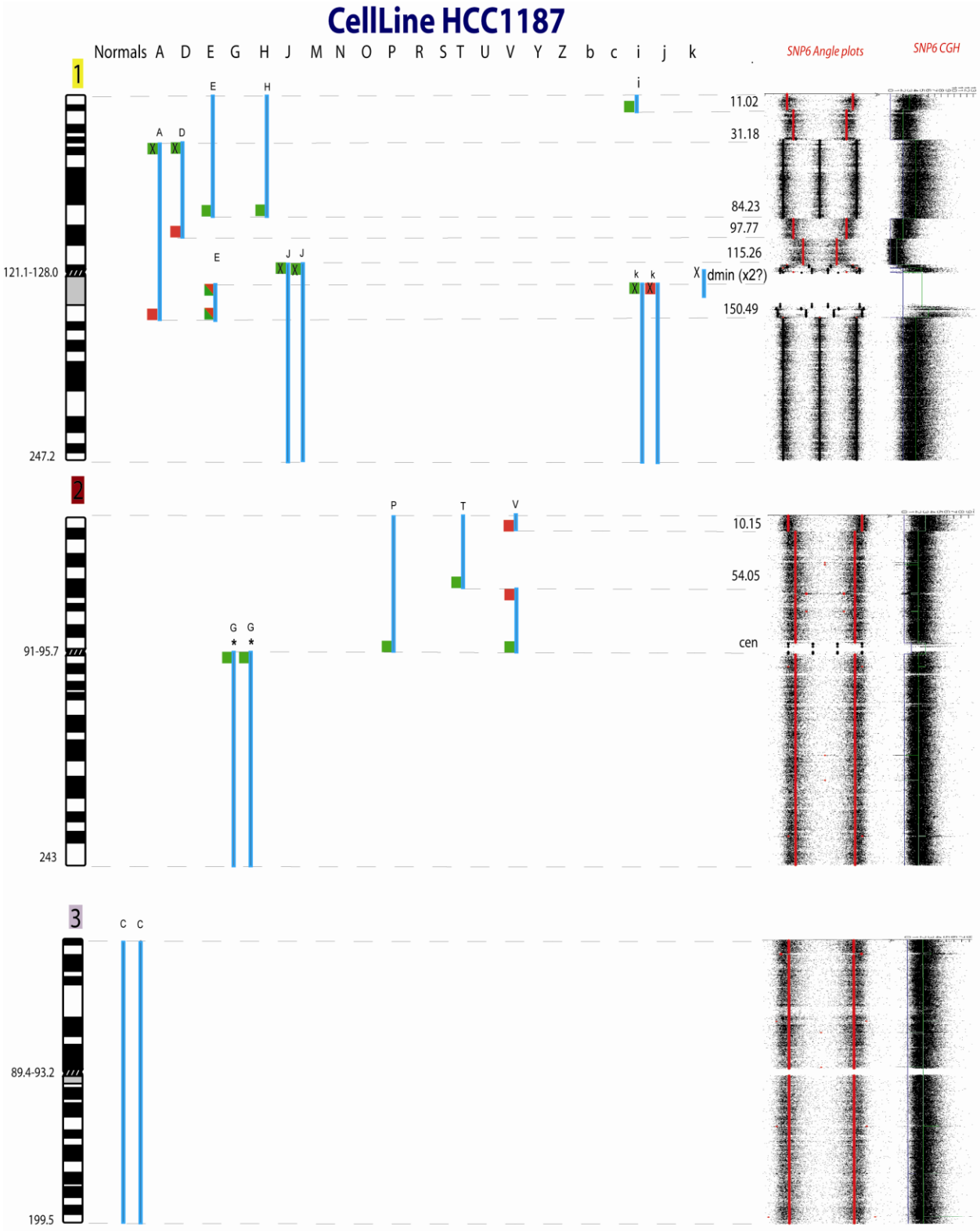
V: Evolution 4: NCI-H209

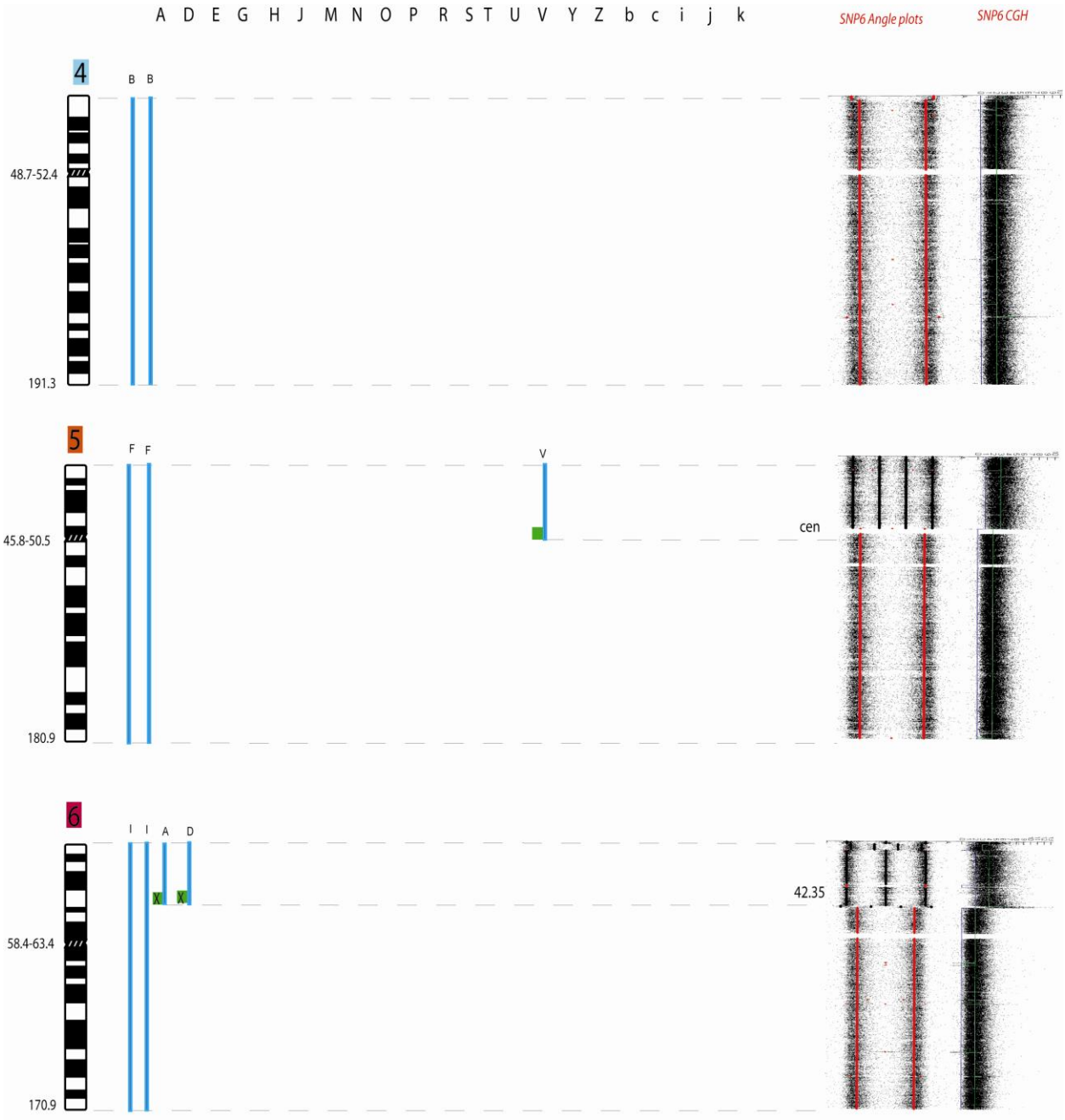

Chr 1
Chr 4
Chr 5
CD4
CD1
UTa
UTb
UTc
INVa
INVb

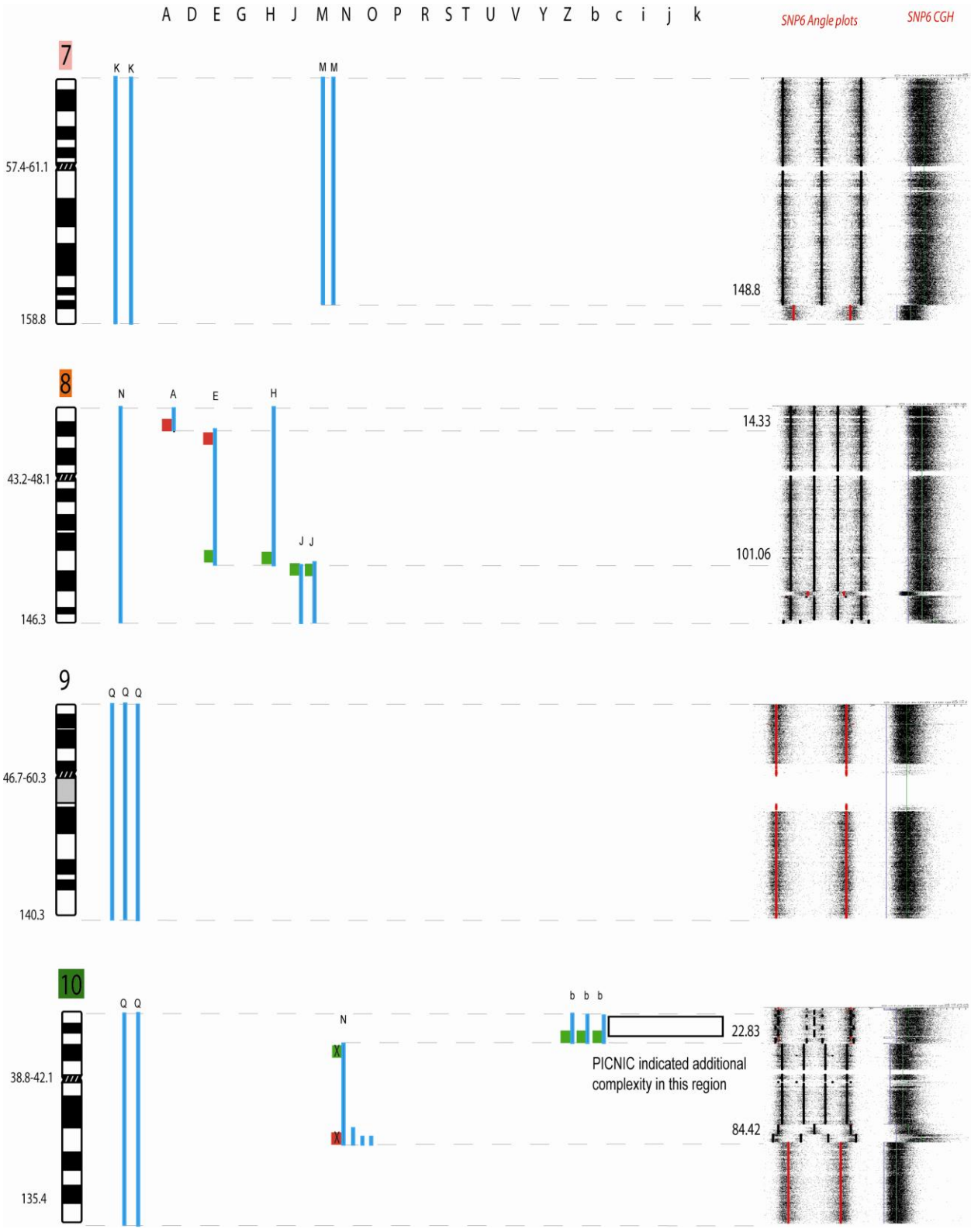


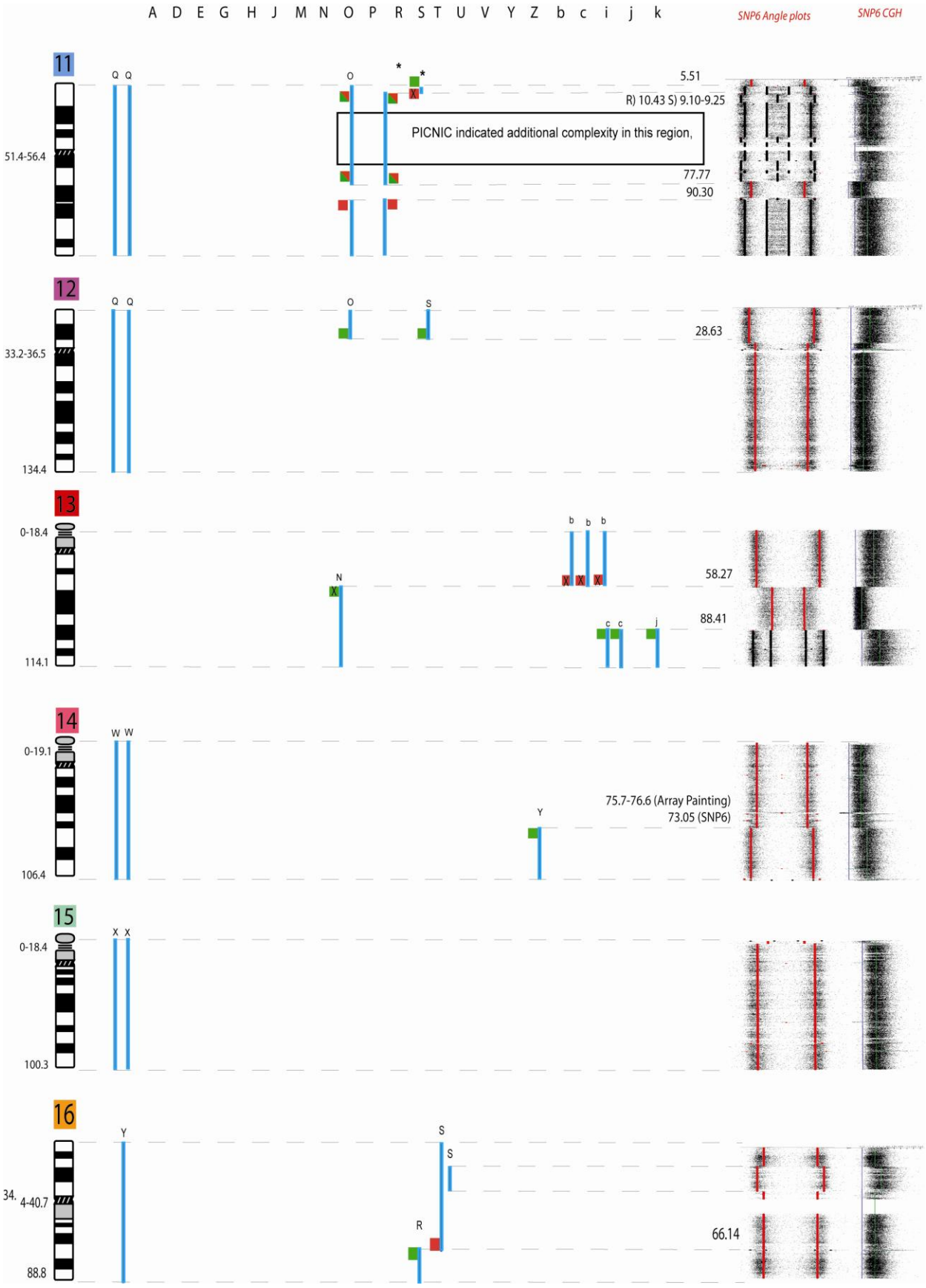
The opaque solution is the alternative position of the deletion. It is only rejected because FISH data suggests an entire chromosome 1 is present.

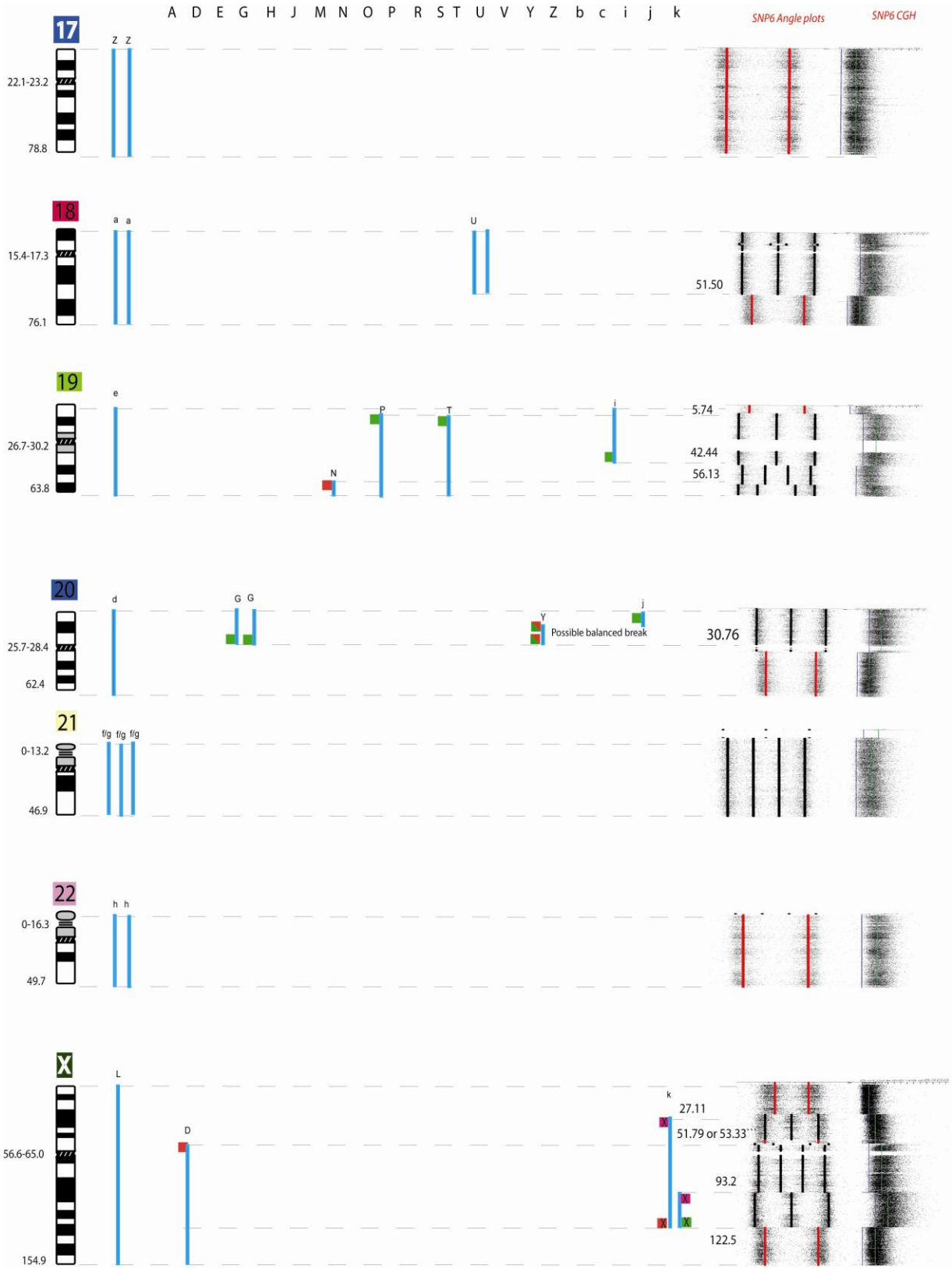
I. HCC1187 Karyotype - Copy Number Comparison



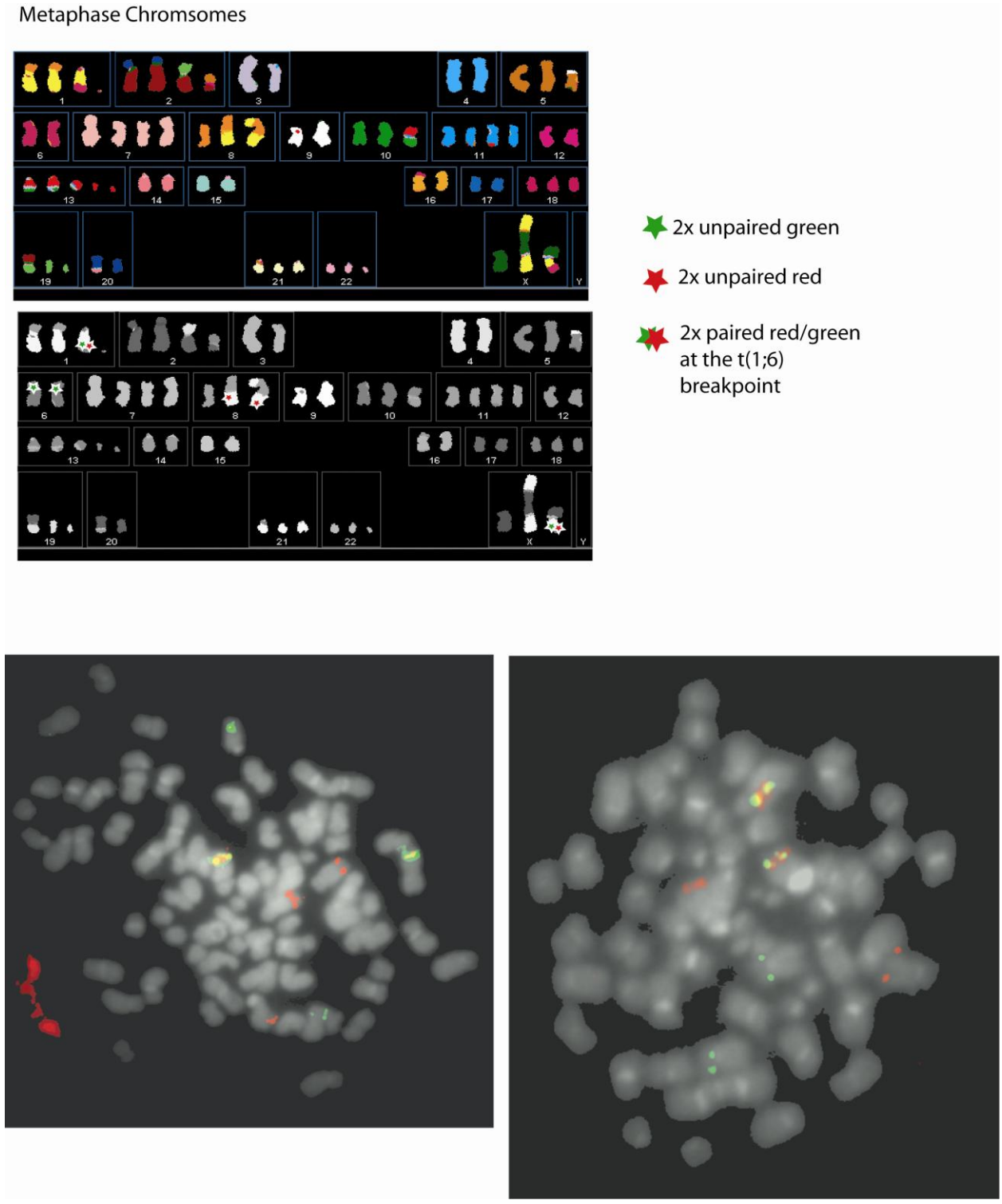

A D E G H J M N O P R S T U V Y Z b c i j k
SNP6 Angle plots
SNP6 CGH
4
B B
48.7-52.4
191.3
5
F F
V
45.8-50.5
180.9
cen
6
I I A D
58.4-63.4
170.9
42.35


A D E G H J M N O P R S T U V Y Z b c i j k
SNP6 Angle plots
SNP6 CGH
7
K K
M M
57.4-61.1
158.8
148.8
8
N
A
E
H
J J
43.2-48.1
146.3
14.33
101.06
9
Q Q Q
46.7-60.3
140.3
10
Q Q
N
b b b
38.8-42.1
135.4
22.83
84.42
PICNIC indicated additional complexity in this region


A D E G H J M N O P R S T U V Y Z b c i j k
SNP6 Angle plots
SNP6 CGH
11
51.4-56.4
Q Q
O
*
*
X
PICNIC indicated additional complexity in this region,
5.51
R) 10.43 S) 9.10-9.25
77.77
90.30
12
33.2-36.5
134.4
Q Q
O
S
28.63
13
0-18.4
114.1
N
X
b b b
X X X
c c
j
58.27
88.41
14
0-19.1
106.4
W W
Y
75.7-76.6 (Array Painting)
73.05 (SNP6)
15
0-18.4
100.3
X X
16
34.4-40.7
88.8
Y
S
S
R
66.14

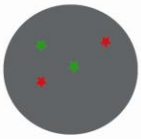

A D E G H J M N O P R S T U V Y Z b c i j k
SNP6 Angle plots
SNP6 CGH
17
22.1-23.2
78.8
Z Z
18
15.4-17.3
76.1
a a
U
51.50
19
26.7-30.2
63.8
e
N
P
T
i
5.74
42.44
56.13
20
25.7-28.4
62.4
d
G G
Y
Possible balanced break
j
30.76
21
0-13.2
46.9
f/g f/g f/g
22
0-16.3
49.7
h h
X
56.6-65.0
154.9
L
D
k
27.11
51.79 or 53.33
93.2
122.5

II. HCC1187 FISH



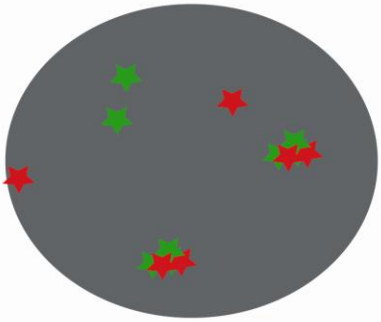
Interphase nuclei

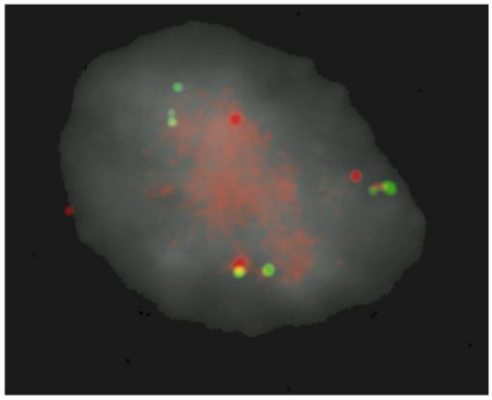
Control



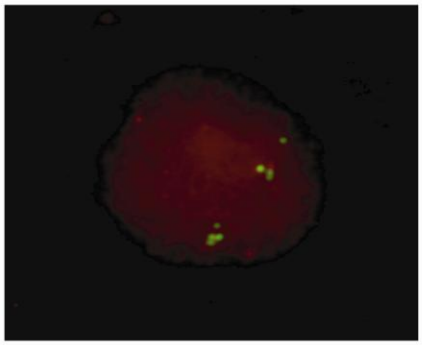
2 unpaired reds
2 unpaired greens

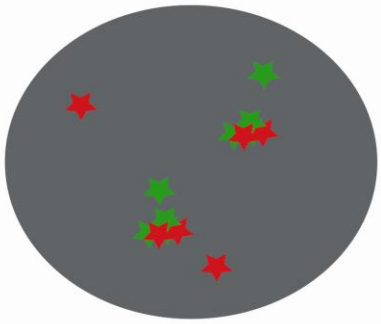
HCC1187

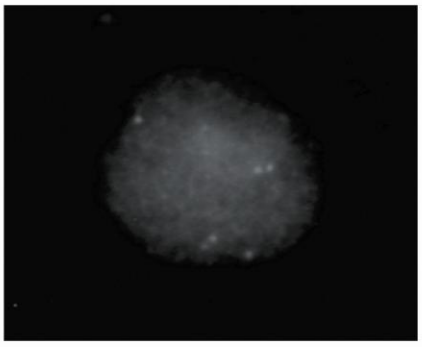


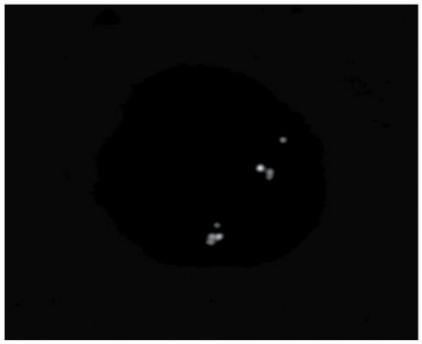


2 unpaired reds
2 unpaired greens
2 clusters of fused signals







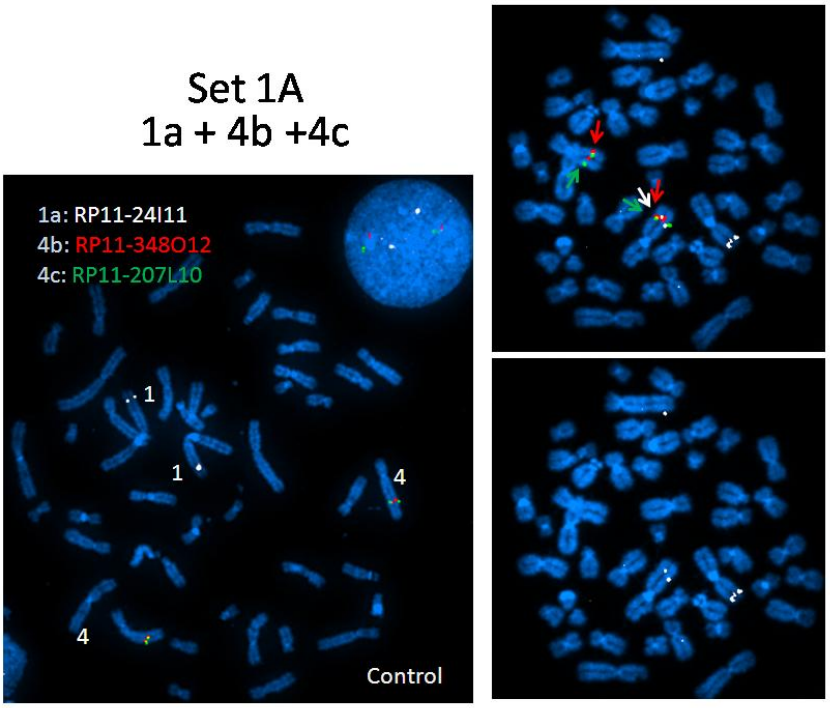


II. NCI-H209

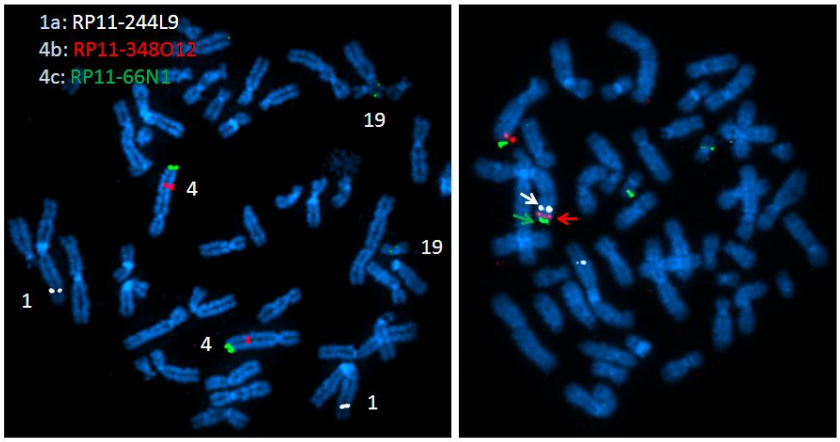
List of BAC clones ordered for use in FISH

Chr 1	Clone ID	GRCh37/hg19	NCBI36/hg18	Size	Code
1a	RP11-24I11	29,142,364-29,315,591	29,014,951-29,188,178	173,228	
	RP11-5B9	29,349,372-29,505,615	29,221,959-29,378,202	156,244	
	RP11-244L9	29,705,134-29,871,544	29,577,721-29,744,131	166,411	
1b	RP11-3B21	30,500,135-30,668,180	30,272,722-30,440,767	168,046	
	RP11-158E15	35,653,873-35,827-255	35,426,460-35,599,842	173,383	
	RP11-204N21	38,543,463-38,704,645	38,316,050-38,477,232	161,183	
Chr 4					
4a	RP11-279O6	132,073,200-123,223,792	132,292,650-132,443,242	150,593	
	RP11-103E19	133,156,783-133,338,920	133,376,233-133,558,370	182,138	
	RP11-39B24	133,651,899-133,825,512	133,871,349-134,044,962	173,614	
4b	RP11-348O12	134,115,077-134,258,629	134,334,527-134,478,079	143,553	
4c	RP11-207L10	134,561,477-134,705,934	134,780,927-134,925,384	144,458	
	RP11-314C18	155,499,038-155,685,533	155,718,488-155,904,983	186,496	
	RP11-66N1	188,580,804-188,709,319	188,817,798-188,946,313	128,516	
Chr 5					
5a	RP11-118N3	95,420,703-95,580,383	95,446,459-95,606,139	159,681	
	RP11-58B2	101,719,310-101,881,680	101,747,209-101,909,579	162,371	
	RP11-167O18	105,649,017-105,787,508	105,676,916-105,815,407	138,492	

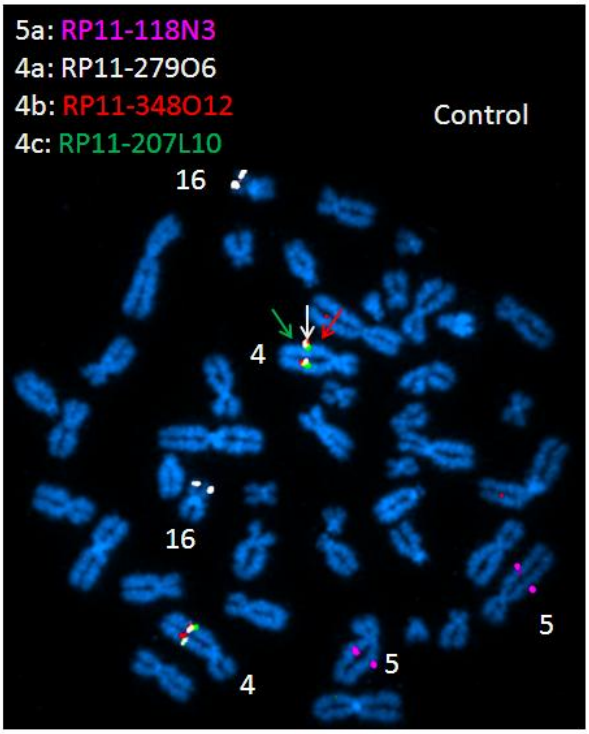
Note that that shade clones have not been subject to FISH experiment

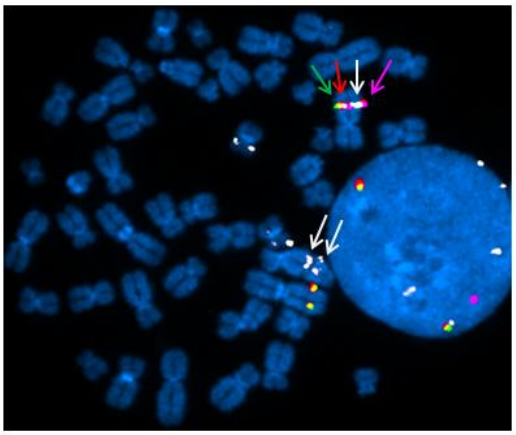


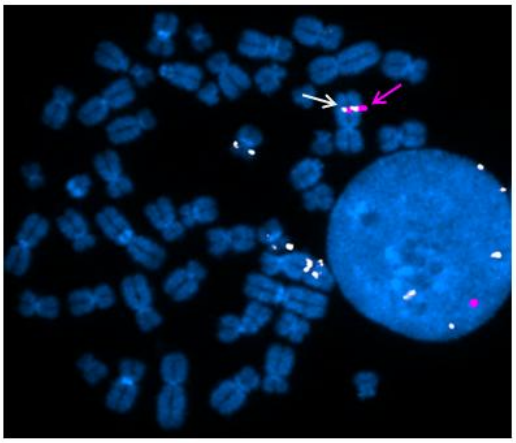
Set 1B: 1a + 4b + 4c



Set 2A
5a+ 4a+4b+4c

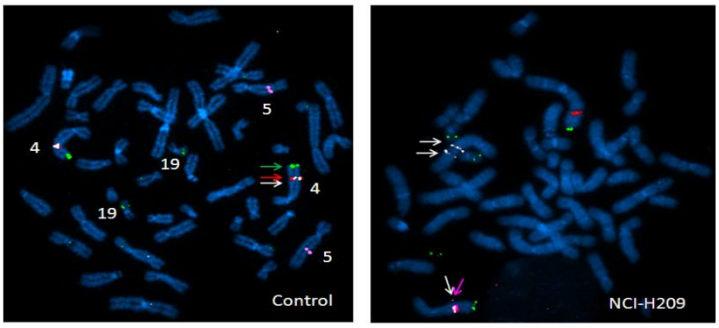






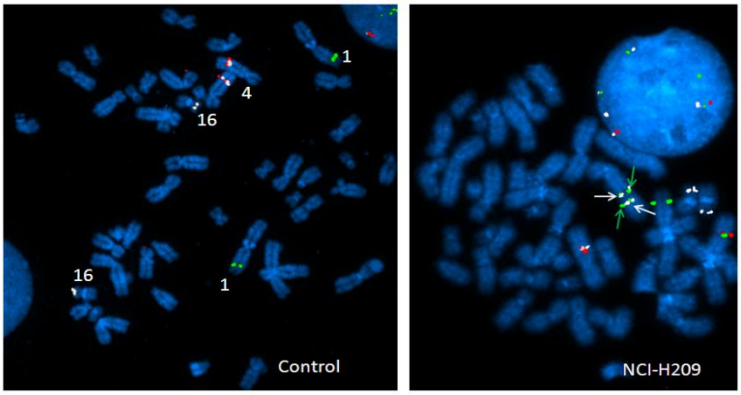
Set 2B: 5a+ 4a+4b+4c

5a: RP11-167O18; 4a: RP11-39B24 4b; RP11-348O12; 4c: RP11-66N21



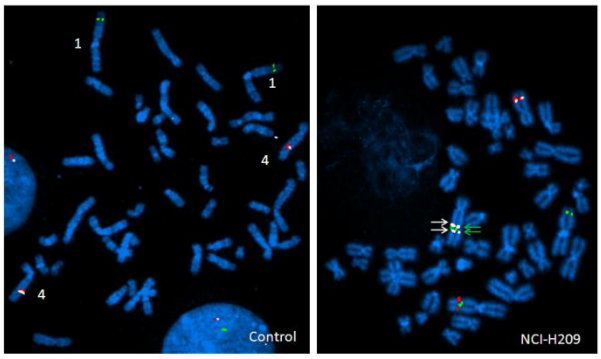
Set 3A: 4a+4b+1b

4a: RP11-27906; 4b: RP11-348O12; 1b: RP11-3B21



Set 3B: 4a+4b+1b

4a: RP11-39B24; 4b: RP11-348O12; 1b: RP11-204N21



7. Segments and Rearrangements Table

		Position			Copy Number			Mutation Multiplicity			
Chromosome	Segment	Start	Finish	Length	Total	Minor	Major	x1	x2	x3	x4
17	1	0.56	20.03	19.47	1	0	1	17	na	na	na
17	2	20.09	36.75	16.66	2	1	1	36	na	na	na
17	3	36.75	81.05	44.30	3	1	2	128	14	na	na
6	1	0.56	24.74	24.17	2	0	2	64	3	na	na
6	2	24.74	65.17	40.43	1	0	1	32	na	na	na
5	1	0.56	53.67	53.11	2	1	1	112	na	na	na
5	2	53.67	140.07	86.40	1	0	1	67	na	na	na
13	1	19.43	54.97	35.55	1	0	1	21	na	na	na
13	2	54.97	55.97	1.00	3	1	2	0	0	na	na
13	3	55.97	55.97	<0.01	3	1	2	0	0	na	na
13	4	55.97	58.58	2.61	1	0	1	3	na	na	na
13	5	58.59	59.59	1.00	3	1	2	2	1	na	na
13	6	59.59	60.16	0.57	1	0	1	0	na	na	na
13	7	60.17	60.92	0.75	3	1	2	0	1	na	na
13	8	60.92	67.76	6.84	1	0	1	11	na	na	na
13	9	67.78	67.79	0.01	3	1	2	62	12	na	na
13	10	67.79	95.89	28.10	3	1	2	0	0	na	na
13	11	95.89	115.11	19.21	2	1	1	42	na	na	na
1	1	11.1	31.2	20.1	2	0	2	-	-	-	-
1	2	31.2	31.3	0.1	6	2	4	-	-	-	-
1	3	31.3	84.2	52.9	4	2	2	-	-	-	-
6	1	5.7	41.4	35.7	4	2	2	-	-	-	-
6	2	41.4	42.4	1	6	2	4	-	-	-	-
6	3	42.4	170.8	127.6	2	0	2	-	-	-	-
1	1	0	29.15	29.15	1	0	1	81	na	na	na
1	2	29.15	30.03	0.88	2	1	1	9	na	na	na
1	3	30.03	41.17	11.14	4	1	3	255	21	4	na
1	4	41.17	42.38	1.21	2	1	1	17	na	na	na
1	5	42.38	42.58	0.2	4	1	3	8	0	0	na
1	6	42.58	44.26	1.68	3	1	2	25	1	na	na
1	7	44.26	44.28	0.02	5	1	4	0	0	0	0
1	8	44.28	48.58	4.3	1	0	1	12	na	na	na
1	9	48.58	51.28	2.7	3	1	2	23	8	na	na
1	10	51.28	54.47	3.19	1	0	1	7	na	na	na
1	11	54.47	246.01	191.54	2	1	1	2362	na	na	na
4	1	114.58	131.81	17.23	2	0	2	241	70	na	na
4	2	131.81	134.09	2.28	3	1	2	42	33	na	na
4	3	134.09	134.23	0.14	4	1	3	5	1	1	na
4	4	134.23	191.26	57.03	2	1	1	1303	na	na	na
13	1	0	92.2	92.2	2	0	2	192	331	na	na
17	1	0	75.3	75.3	2	0	2	153	127	na	na

8. Genome Graphs