Supplementary Material

Phylogenetic analysis of multiprobe fluorescence in situ hybridization data from tumor cell populations

Salim Akhter Chowdhury[1,2], Stanley E. Shackney[3], Kerstin Heselmeyer-Haddad[4], Thomas Ried[4], Alejandro A. Schäffer[5], Russell Schwartz[2,6]

[1]Joint Carnegie Mellon/University of Pittsburgh Ph.D. Program in Computational Biology and [2]Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA, USA, [3]Intelligent Oncotherapeutics, Pittsburgh, PA, USA, [4]Genetics Branch, Center for Cancer Research, NCI, NIH, Bethesda, MD, USA, [5]Computational Biology Branch, NCBI, NIH, Bethesda, MD, USA, [6]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

SS1: Classification Performance

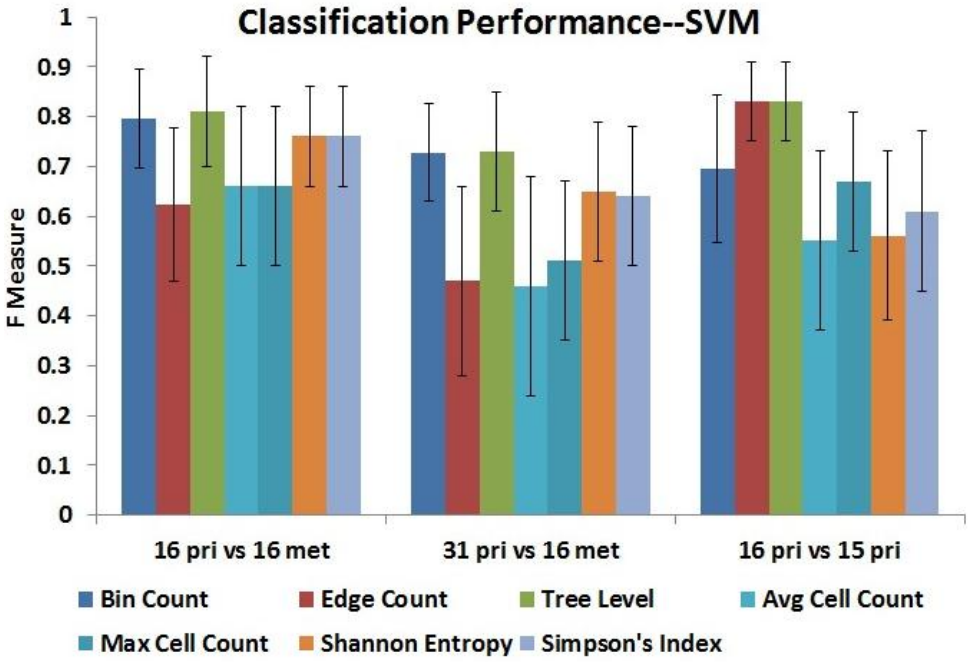

Figure SF1: Performance of tree-based versus cell-based features in classification tasks using an SVM classifier. Each chart shows accuracy of three tree-based and four cell-based feature sets on the three defined prediction tasks reported using F measure, the harmonic mean of precision and recall.

SS2: Simulation tests

In order to validate accuracy of tree inferences, we simulated progression trees by the following protocol, which takes a random number $r \in [0,1]$, a number gene probes Np to be simulated and the number of cells K to be sampled:

(1) Fix the root node to be $R = \{2\}^{Np}$
(2) Initialize the tree $T = R$ and depth of the tree $d = 1$
(3) Extract all nodes N generated at depth d
(4) For each node $n \in N$, repeat the following
  (a) Generate a uniform random number $U \in [0,1]$
  (b) if U > r, then do the following:
    (i) Select a gene probe q from the copy number profile of n randomly
    (ii) Select a direction gl of unit copy number change (gain or loss) from the copy number profile of n randomly
    (iii) Generate a child C of n with copy number profile based on q and gl
    (iv) If the copy number profile of C has not been generated previously then put C into T
    (v) Go to step 4(a)
(5) If no new child node is generated at this iteration then stop, otherwise increment d and go to step (3)
(6) For each $k \in K$, generate a uniform integer random number I_R between 1 and |T| and assign k to the copy number profile of the node indexed by I_R.

Simulations were conducted for the present work with parameters $r = 0.5$, $Np = 4$ and $K = 250$. In order to compare the similarity between the simulated tree T_S and the inferred tree T_I, we used a weighted matching based metric. We first identified the set of non-trivial bipartitions in both simulated and inferred tree, represented by $\prod(T_S)$ and $\prod(T_I)$ respectively. Then we defined a complete weighted bipartite graph $G(A,B,E)$, where A and B consist of $\prod(T_S)$ and $\prod(T_I)$ respectively. Since T_S and T_I can have non disjoint node sets, we calculated the weight between two vertices of G by dividing the cardinality of shared nodes set between the bipartitions represented by those two vertices with the total number of nodes in the simulated tree. Then the distance between T_S and T_I is calculated by identifying the maximum weight matching in $G(A,B,E)$. After dividing the weight of the matching by the total number of nodes in T_S, we calculated the fractional similarity in bipartitions between the two trees. We calculated the matching accuracy by multiplying the fractional similarity with 100.

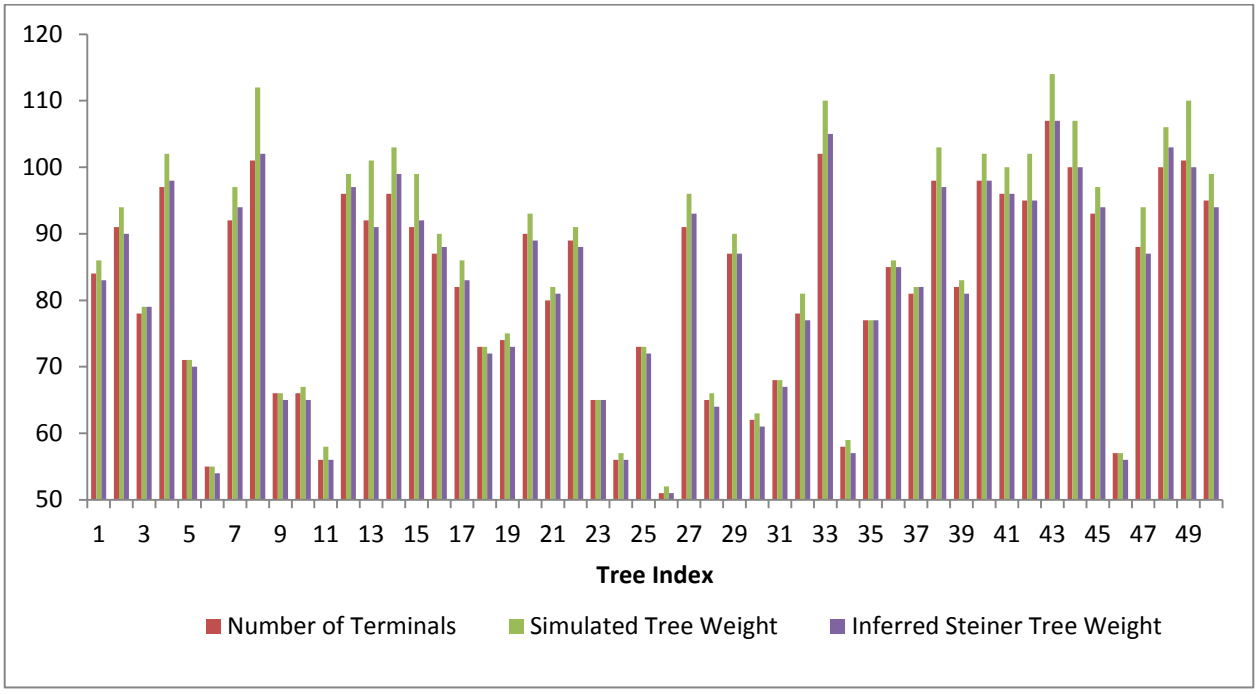

Figure SF2: Comparison of simulated and inferred Steiner tree weights for fifty simulated trees. For each tree, the total number of terminals, real tree weight and inferred Steiner tree weight are shown.

We now summarize the results from our simulation test. We generated the simulated trees using the procedure described above. We picked the terminal nodes (those with at least one cell) and used our approximation algorithm to infer the tumor progression tree. To generate a dataset comparable to our real data, we selected only simulated trees with total number of terminals in the range of 50 and 120. A comparison of the 50 trees is shown Figure SF1. For each case, Figure SF1 shows the total number of terminals and the weights of the simulated and inferred tree. We can see from the comparison that the weight of the inferred tree is generally close to that of the simulated tree but never exceeds it. This observation would suggest that the heuristic is effective in finding parsimonious trees but that the parsimony model leads to a small bias towards underestimating true tree cost.

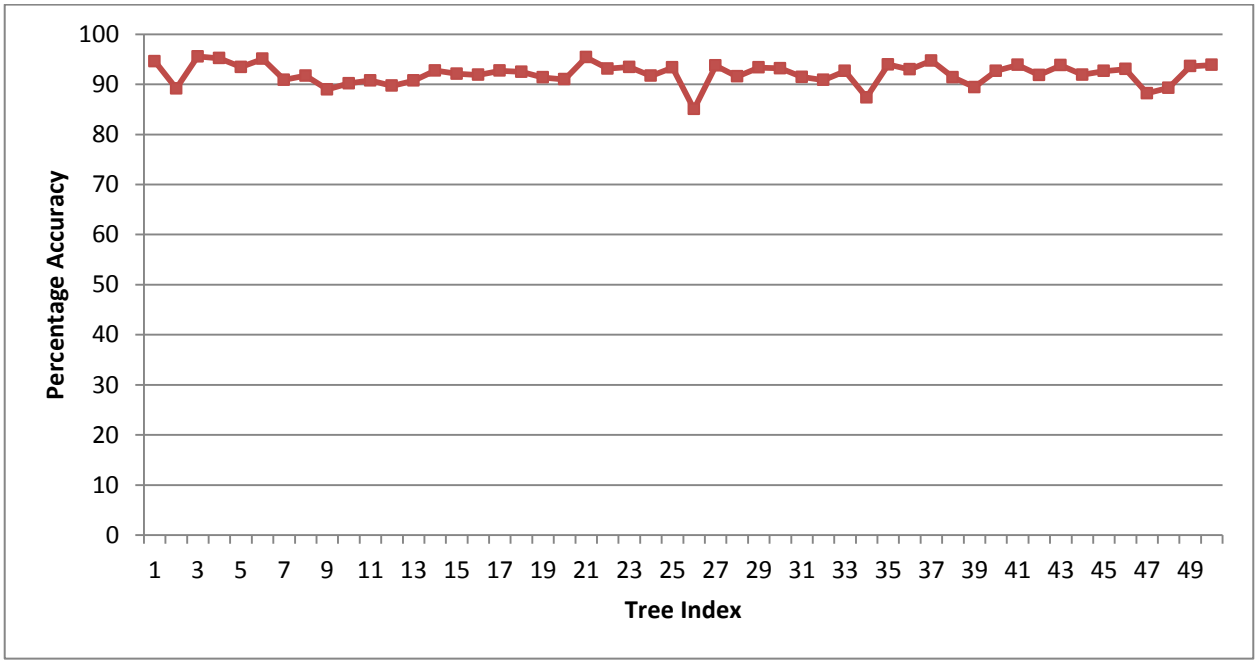

Figure SF3: Percentage accuracy of the set of bipartitions for each inferred Steiner tree with respect to the bipartitions in the corresponding simulated tree.

We show the percentage matching accuracy for each of the 50 test in Figure SF2. The high accuracy for all the cases indicates that the simulated and inferred trees are topologically similar. The mean and standard deviation of the percentage matching accuracies across all fifty examples are 92% and 2.13% respectively. In order to statistically quantify the accuracies, for each simulation case, we generated 100 random trees by randomly picking edges among the nodes in the terminal set and repeated the analysis procedure above to calculate the percentage matching accuracy. We then calculated the p values for the inferred tree matching accuracy by scoring it on the distribution generated by the random trees. For 80% of the cases, the p value is statistically significant. Some of the discrepancies might result from the fact that our algorithm does not generate any Steiner node with degree 2, as these Steiner nodes do not decrease the weights of the maximum parsimony progression trees.

SS3: Comparison of exact and heuristic algorithms

			Exact Approach		Heuristic Approach		
Case	Total Probes	Number of Terminal Nodes	RSMT Weight	Total Runtime in Seconds	RSMT Weight	Total Runtime in Seconds	Percentage of times Call to MST generation routine is avoided in the exact approach
1	3	96	126	910	126	1	80.30
2	3	40	51	10	51	1	85.77
3	3	74	94	318	94	1	67.04
4	3	32	35	1	35	1	99.42
5	3	78	93	150	93	1	63.20
6	3	40	42	1	42	1	96.11
7	3	78	91	47	91	1	94.12
8	3	42	43	1	43	1	99.99
9	3	70	85	16	85	1	91.04
10	3	37	42	1	42	1	99.47
11	3	62	70	7	70	1	89.06
12	3	33	35	1	35	1	99.94
13	3	68	79	4	81	1	95.60
14	3	31	33	1	33	1	93.53
15	3	73	89	165	89	1	61.68
16	3	107	123	100	124	1	97.75
17	3	39	52	87	52	1	75.62
18	3	10	14	1	14	1	36.02
19	3	72	91	473	91	1	67.93
20	3	65	81	154	83	1	83.17
21	4	87	90	10	90	1	99.88
22	4	57	60	4	60	1	99.19
23	4	69	73	57	73	1	93.25
24	4	61	70	236	71	1	99.15
25	4	65	79	813	79	1	88.34
26	4	59	65	11	66	1	99.55
27	4	63	73	87	74	1	98.75
28	4	29	33	11	33	1	94.32
29	4	31	38	16	38	1	90.82
30	4	63	76	250	76	1	97.10
31	4	33	42	35	42	1	90.64
32	4	58	68	129	68	1	99.98
33	4	21	32	63	32	1	28.57
34	4	62	72	28	72	1	98.81
35	4	50	55	8	55	1	98.79
36	4	118	128	157	128	1	99.98
37	4	48	55	18	56	1	97.17
38	4	80	83	56	83	1	99.98
39	4	70	73	17	73	1	99.98
40	4	76	83	30	83	1	99.21
41	4	63	76	140	76	1	99.74
42	4	32	43	144	43	1	94.79
43	4	39	47	82	47	1	98.66
44	5	39	50	35	52	1	96.12
45	5	29	31	12	32	1	99.98
46	5	25	26	125	26	1	99.98
47	5	25	37	1966	37	1	35.75
48	5	15	18	167	18	1	99.82
49	5	39	46	193	46	1	99.50

Table ST1: Example runs of the exact and heuristic approach showing the total number of probes considered, total terminal nodes in the particular case, RSMT weight, total runtimes of the exact and heuristics approach respectively and percentage of times the MST building routine in Algorithm 1 is not executed resulting in reduced running time of the exact approach.

SS4: Individual bin count analysis

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16
Gain of CCND1	17	9	0	0	0	3	4	0	7	186	1	0	4	0	3	10
Gain of PRKAA1	54	4	1	103	1	83	8	1	7	2	1	1	4	161	3	6
Gain of PROX1	6	7	0	7	7	0	1	0	10	3	2	11	21	0	44	0
Gain of LAMP3	18	2	115	4	17	4	11	0	3	1	82	188	6	0	36	12

(A)

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16
Gain of CCND1	7	6	4	3	0	8	1	1	3	0	4	3	0	11	17	3
Gain of PRKAA1	13	36	4	41	2	28	12	3	9	1	3	6	0	8	4	0
Gain of PROX1	2	2	3	37	18	3	2	8	1	2	1	3	20	3	4	0
Gain of LAMP3	10	3	133	0	9	3	3	71	3	81	14	15	4	5	14	114

(B)

Table ST2: Number of cells in the subtrees rooted at the nodes (termed a "bin" in the main manuscript) directly connected to the root node in the cervical cancer dataset. Data is shown for (A) primary and (B) metastatic stage tumor for bins representing gene gains, with the most populated bin highlighted in each case.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13
Loss of DBC2	14	22	48	103	4	0	2	0	37	0	0	17	12
Loss of CDH1	0	11	3	0	6	128	0	160	4	0	0	105	3
Loss of p53	6	8	1	5	8	0	0	5	1	0	1	2	5
Gain of COX-2	0	0	6	6	0	91	0	8	0	0	2	5	2
Gain of MYC	0	1	0	8	55	0	0	0	1	159	0	0	0
Gain of CCND1	0	1	4	1	0	0	61	0	3	0	167	1	0
Gain of HER-2	10	12	3	0	0	0	0	0	0	1	0	2	25
Gain of ZNF217	129	0	5	2	3	0	0	0	1	0	0	0	0

(A)

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13
Loss of DBC2	145	160	5	8	2	4	67	0	0	1	159	8	8
Loss of CDH1	0	8	2	111	48	9	3	0	89	0	6	21	8
Loss of p53	14	0	9	2	3	6	0	0	0	0	3	21	5
Gain of COX-2	5	0	6	0	0	7	0	93	0	0	0	48	70
Gain of MYC	0	0	0	0	15	0	0	0	0	0	0	2	0
Gain of CCND1	2	0	1	0	0	98	0	0	0	0	12	14	0
Gain of HER-2	6	4	3	2	0	15	0	0	0	0	7	69	2
Gain of ZNF217	3	0	5	3	2	6	0	0	5	0	0	0	4

(B)

Table ST3: Number of cells in the subtrees rooted at the nodes (termed a "bin" in the main manuscript) directly connected to the root node in the 13 Breast cancer patients. Data is shown for (A) DCIS and (B) IDC stage tumor for bins representing gain of oncogenes and loss of tumor suppressor genes, with the most populated bin highlighted in each case.

SS5: Tumor progression trees using samples from a breast cancer patient

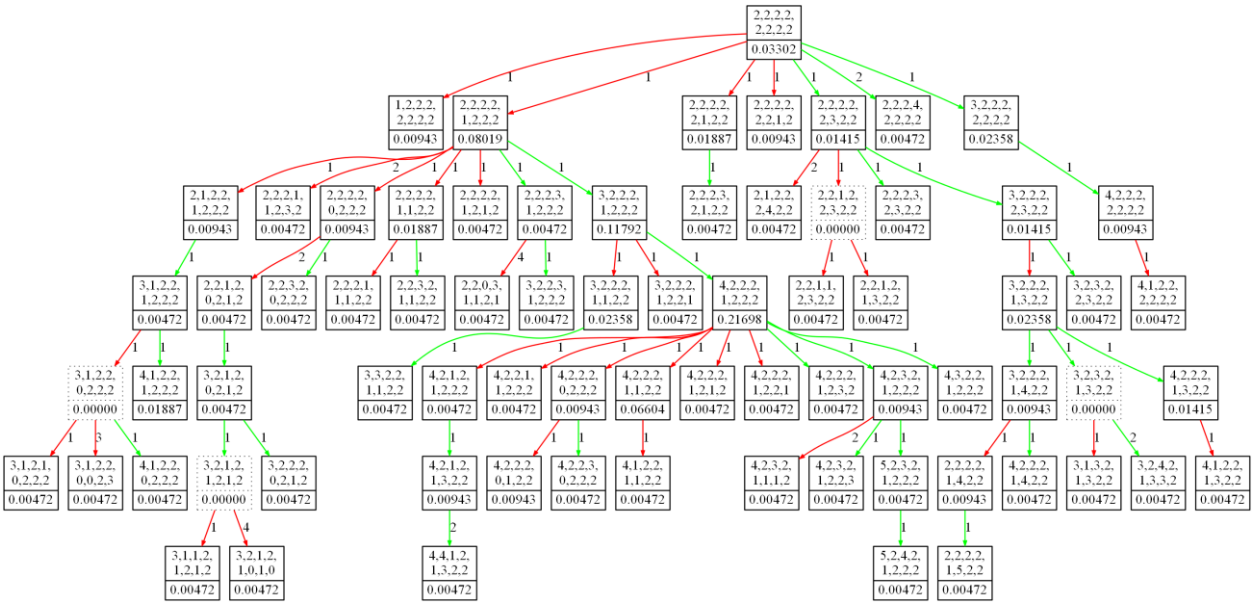

Figure SF4: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 8. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2*, *DBC2*, *MYC*, *CCND1*, *CDH1*, *p53*, *HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes *x* and *y* is the rectilinear distance between the states of *x* and *y*. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

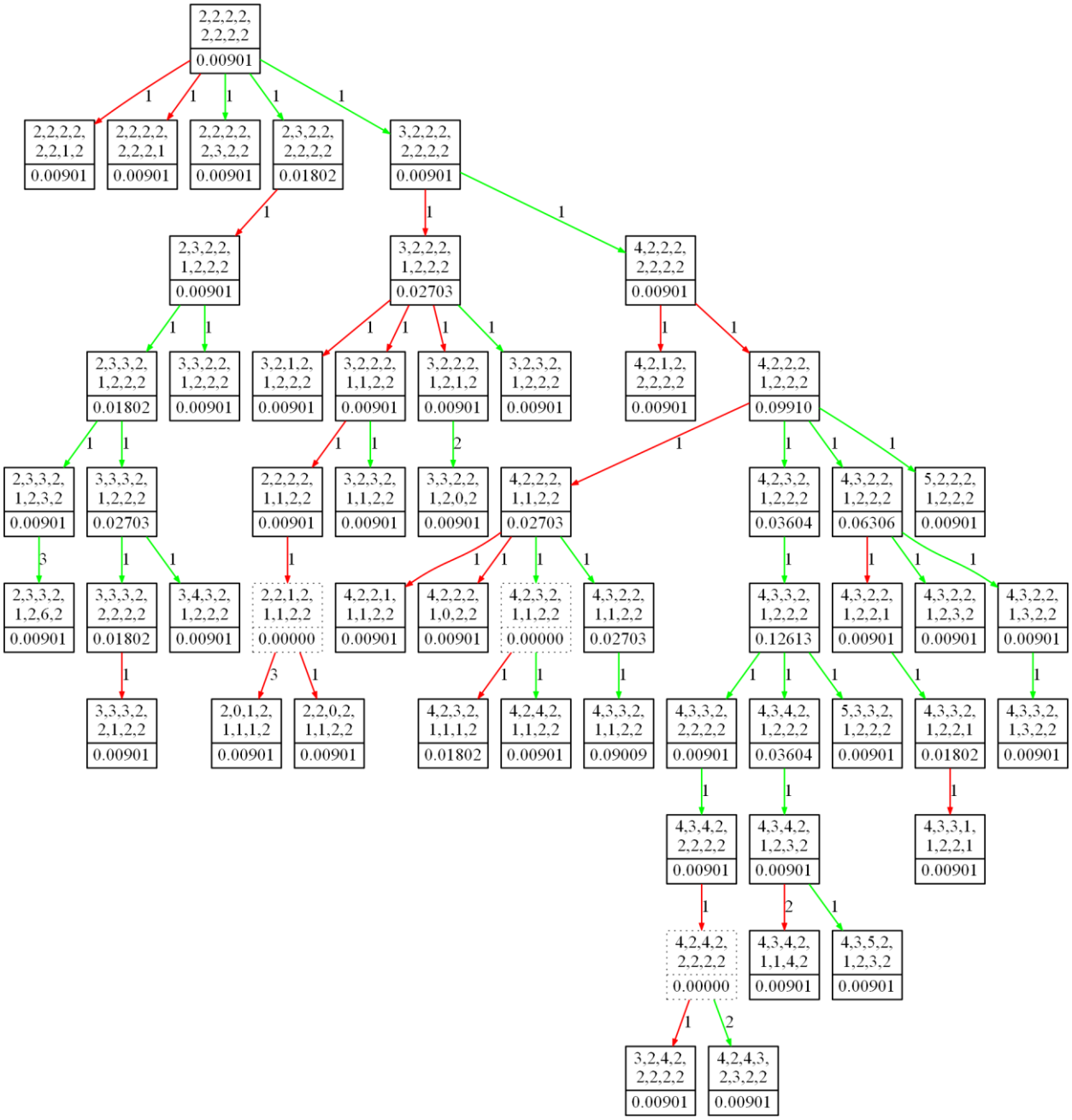

Figure SF5: Phylogenetic tree showing progression of IDC stage breast cancer in patient 8. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2*, *DBC2*, *MYC*, *CCND1*, *CDH1*, *p53*, *HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

SS6: Cervical cancer and Breast cancer tumor progression tree images

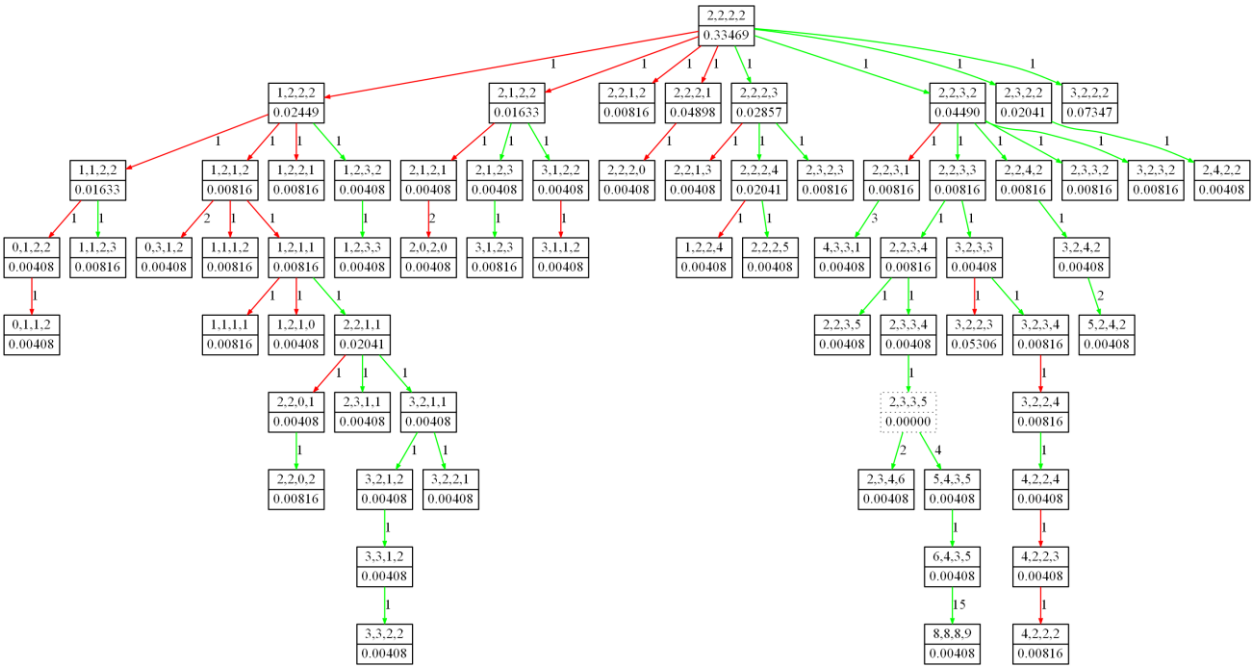

Figure SF6: Phylogenetic tree showing progression of primary stage cervical cancer in patient 1. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3*, *PROX1*, *PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

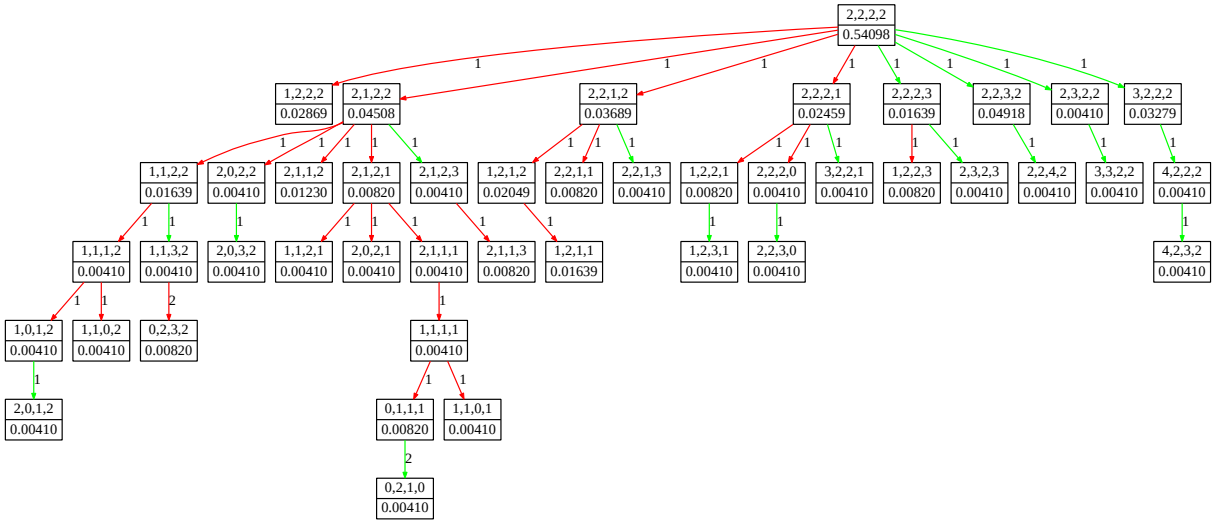

Figure SF7: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 1. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

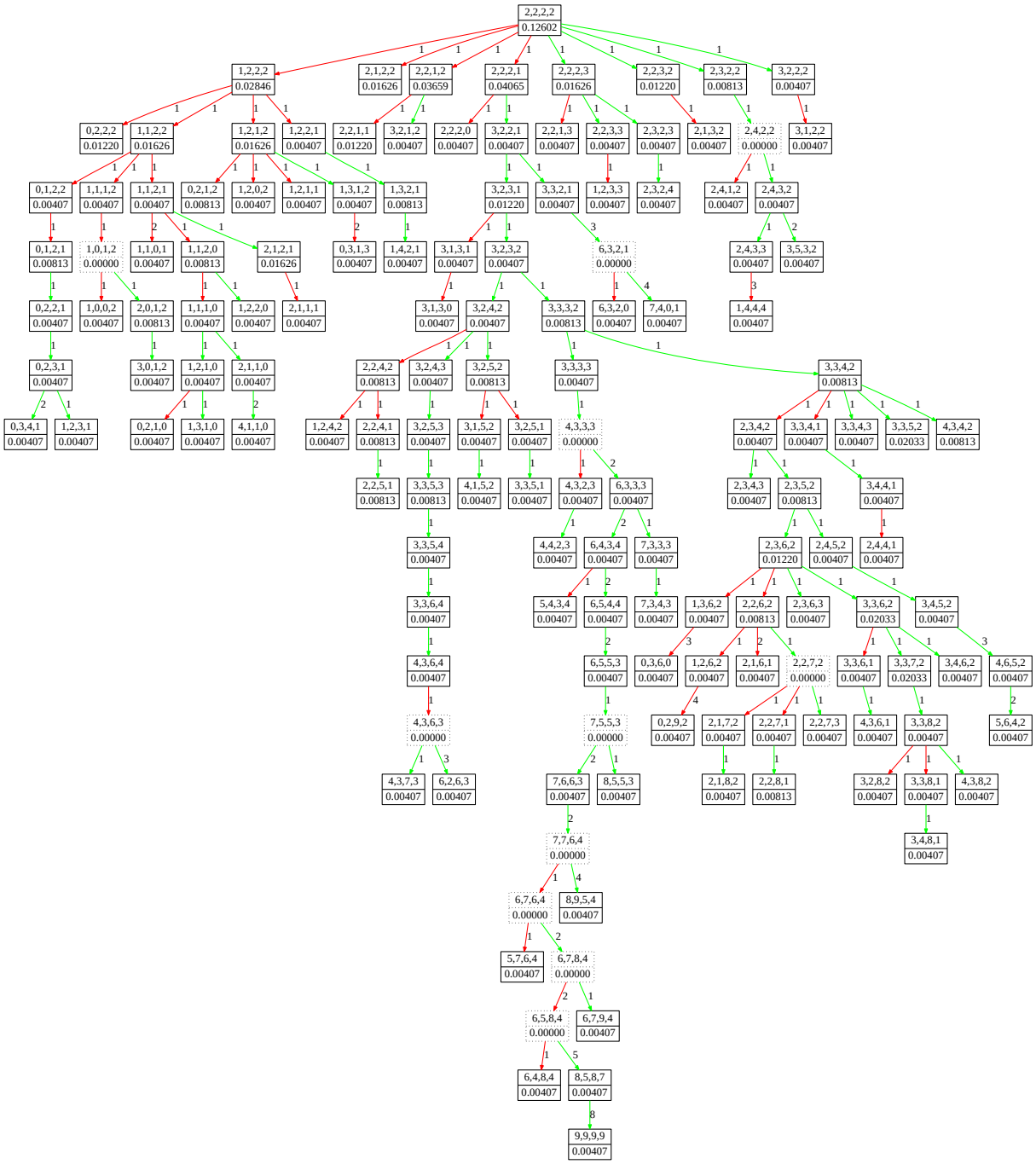

Figure SF8: Phylogenetic trees showing progression of primary stage cervical cancer in patient 2. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

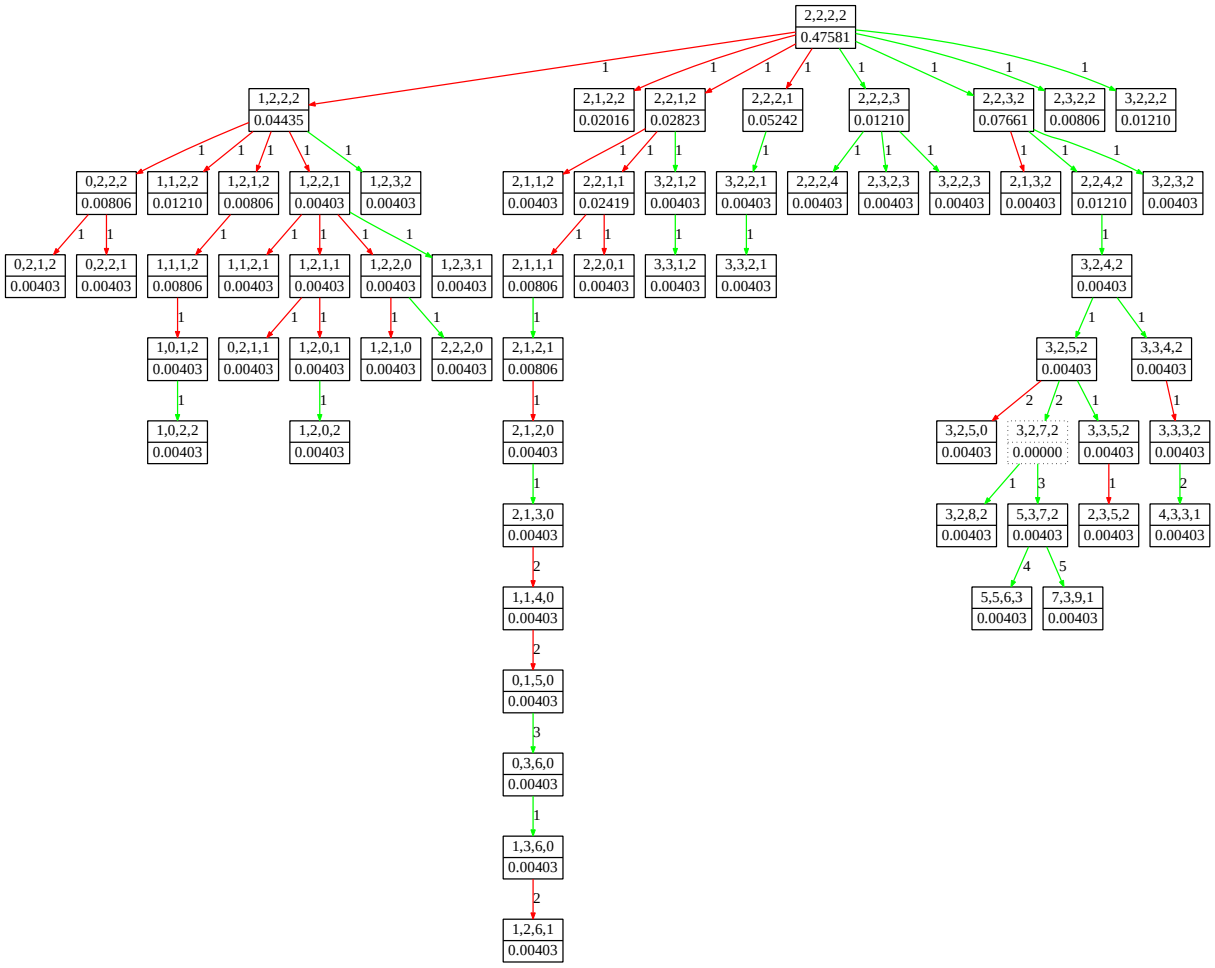

Figure SF9: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 2. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

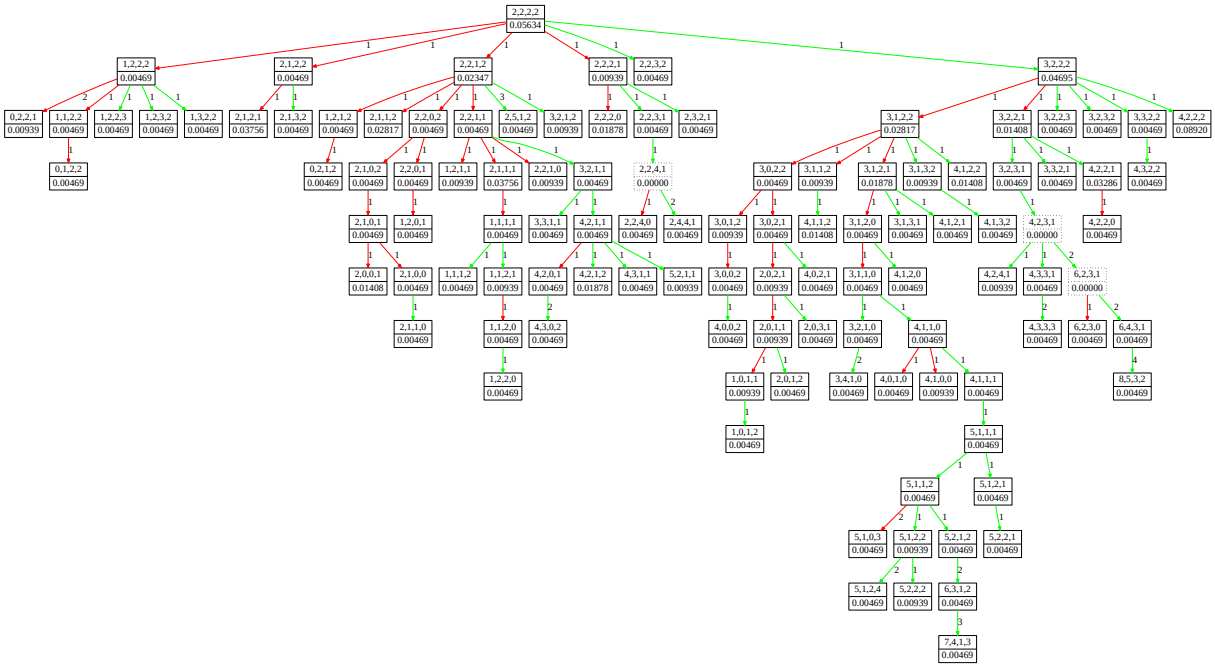

Figure SF10: Phylogenetic trees showing progression of primary stage cervical cancer in patient 3. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

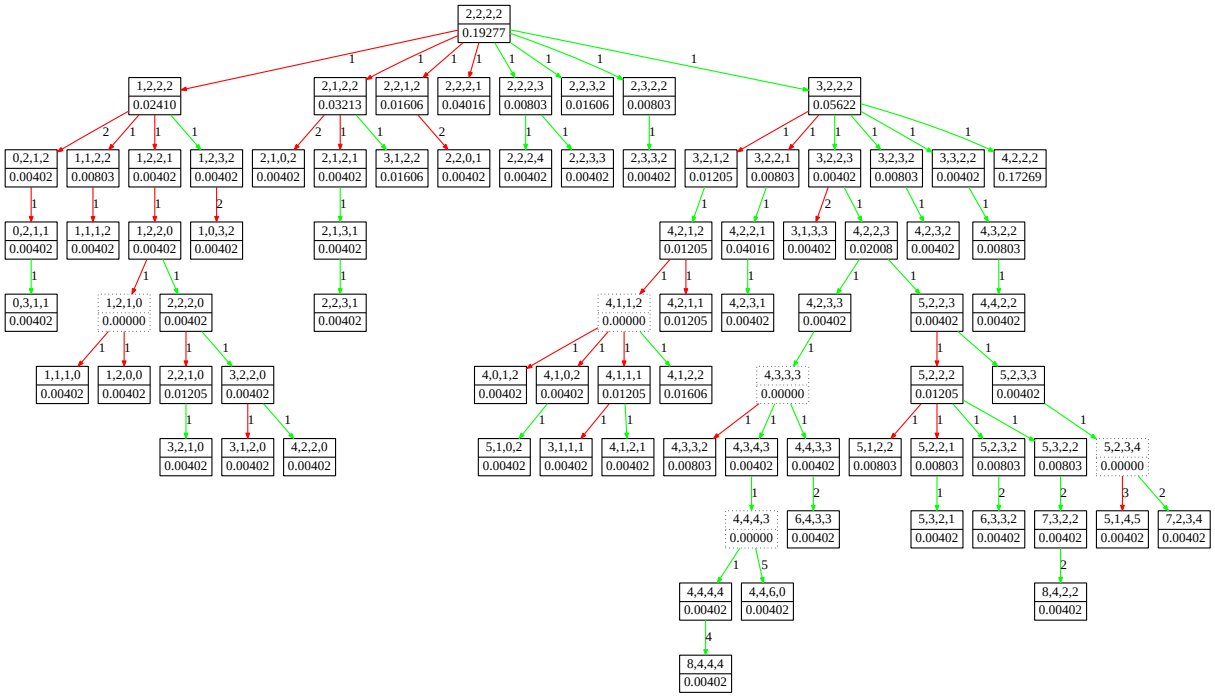

Figure SF11: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 3. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

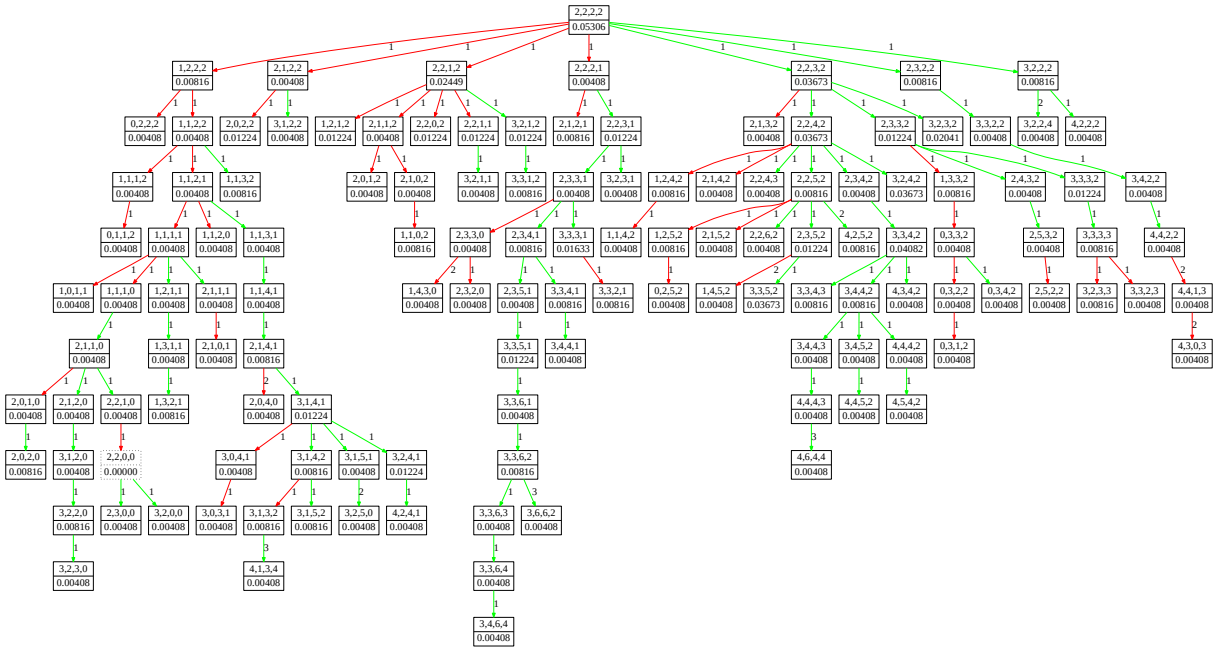

Figure SF12: Phylogenetic trees showing progression of primary stage cervical cancer in patient 4. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

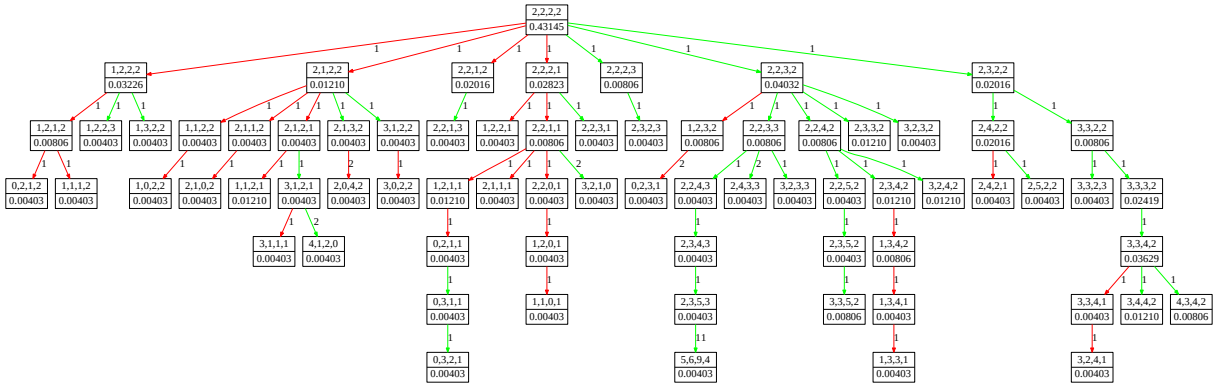

Figure SF13: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 4. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

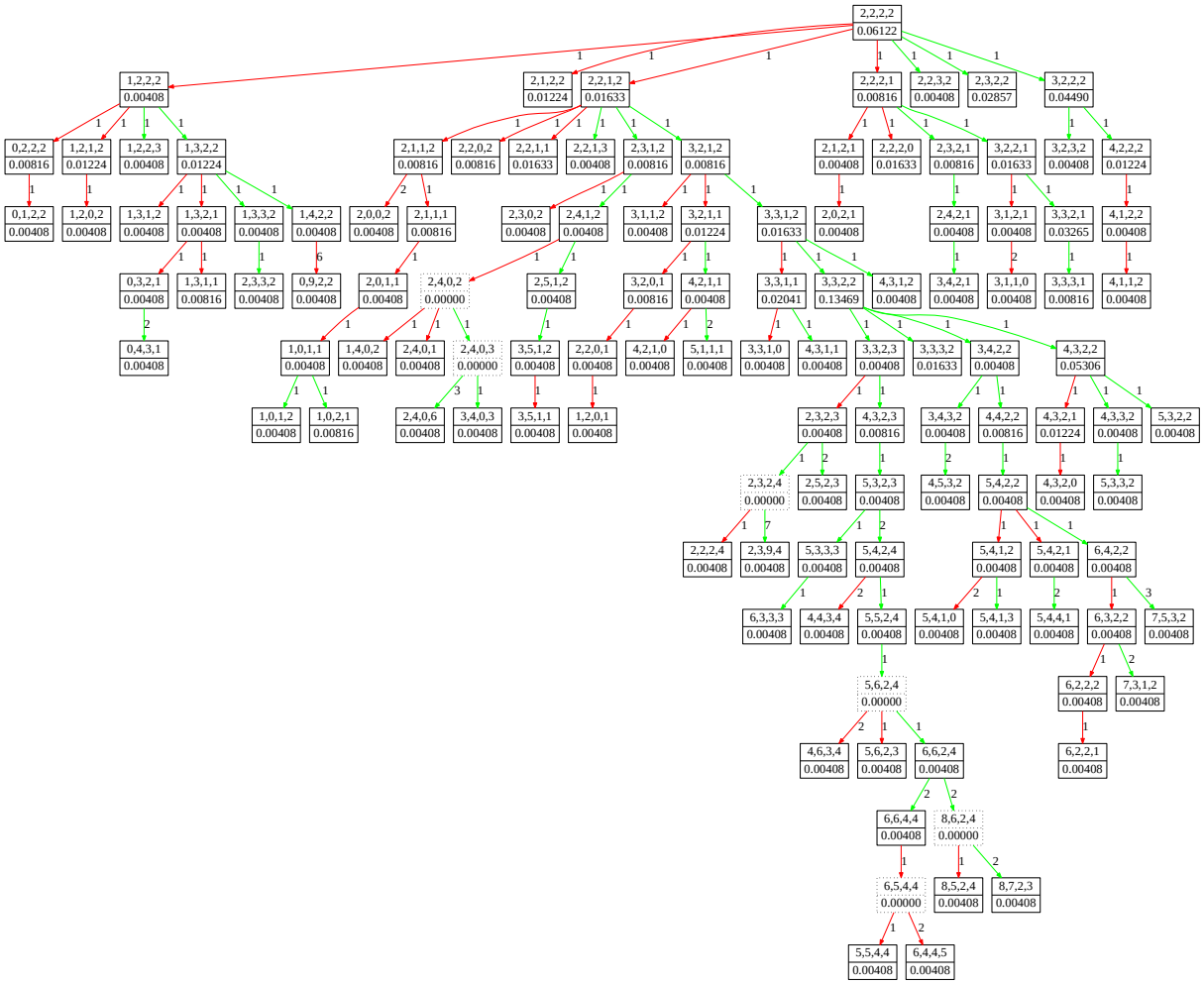

Figure SF14: Phylogenetic trees showing progression of primary stage cervical cancer in patient 5. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

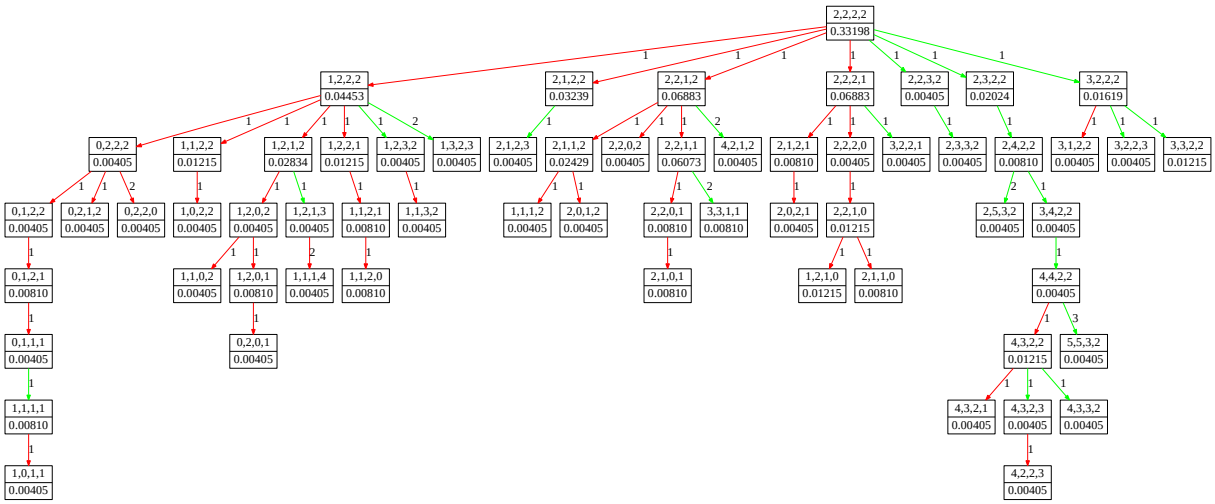

Figure SF15: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 5. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

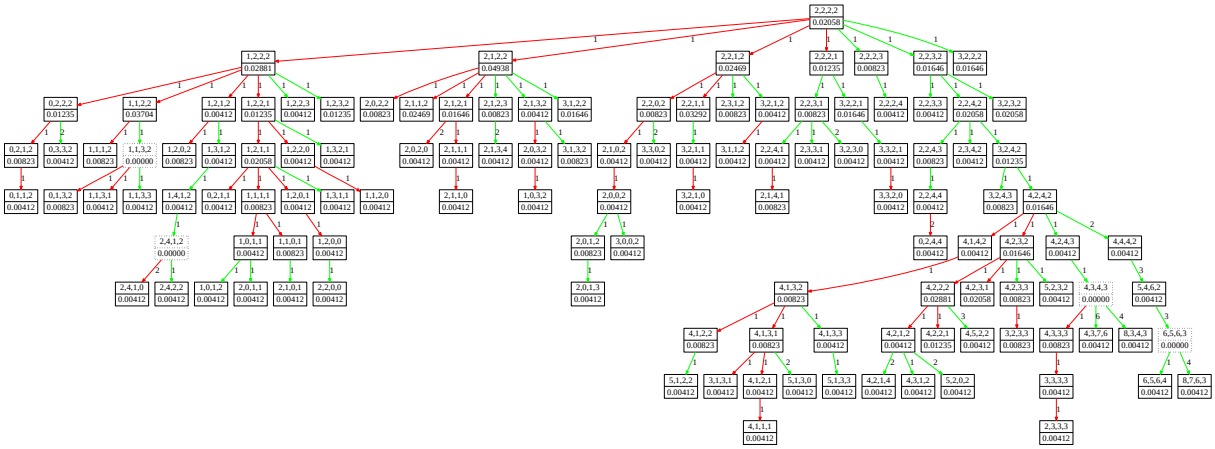

Figure SF16: Phylogenetic trees showing progression of primary stage cervical cancer in patient 6. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

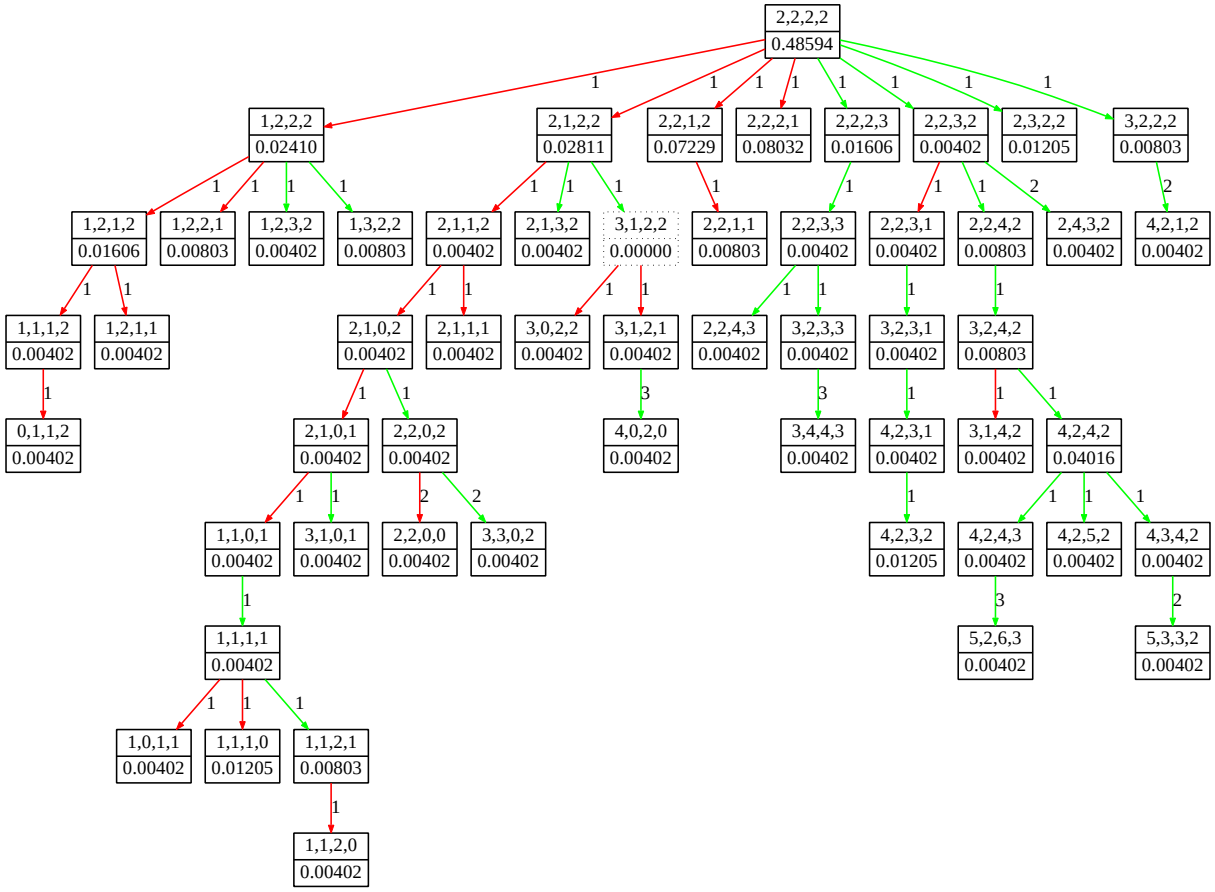

Figure SF17: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 6. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

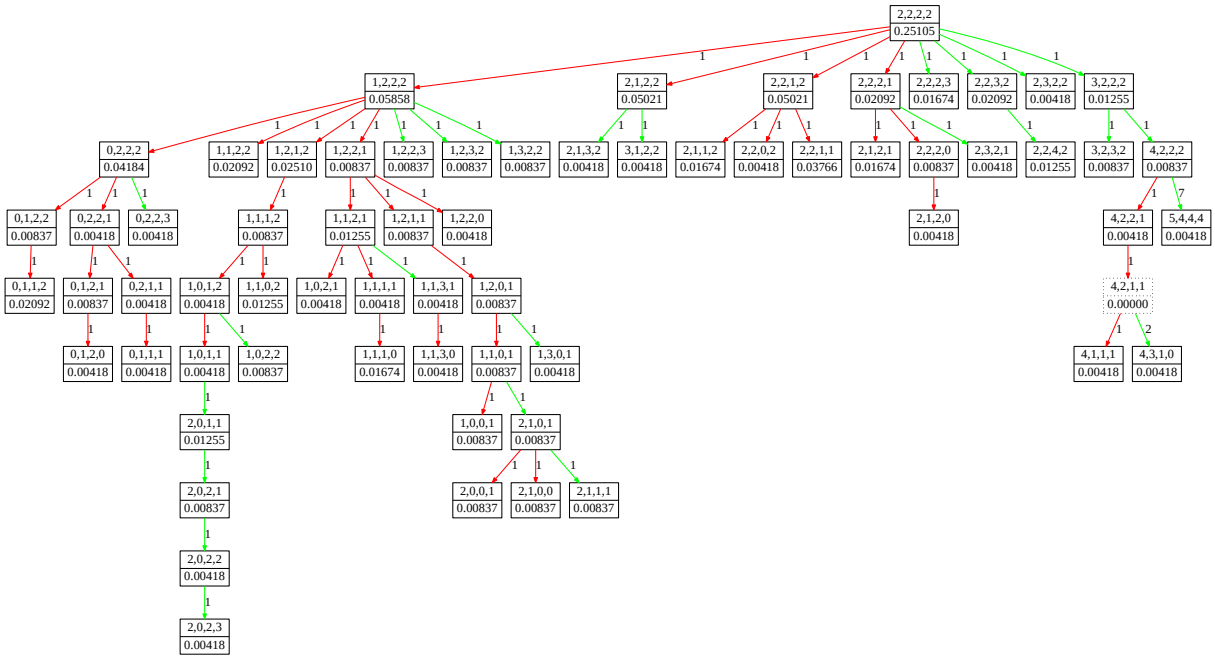

Figure SF18: Phylogenetic trees showing progression of primary stage cervical cancer in patient 7. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

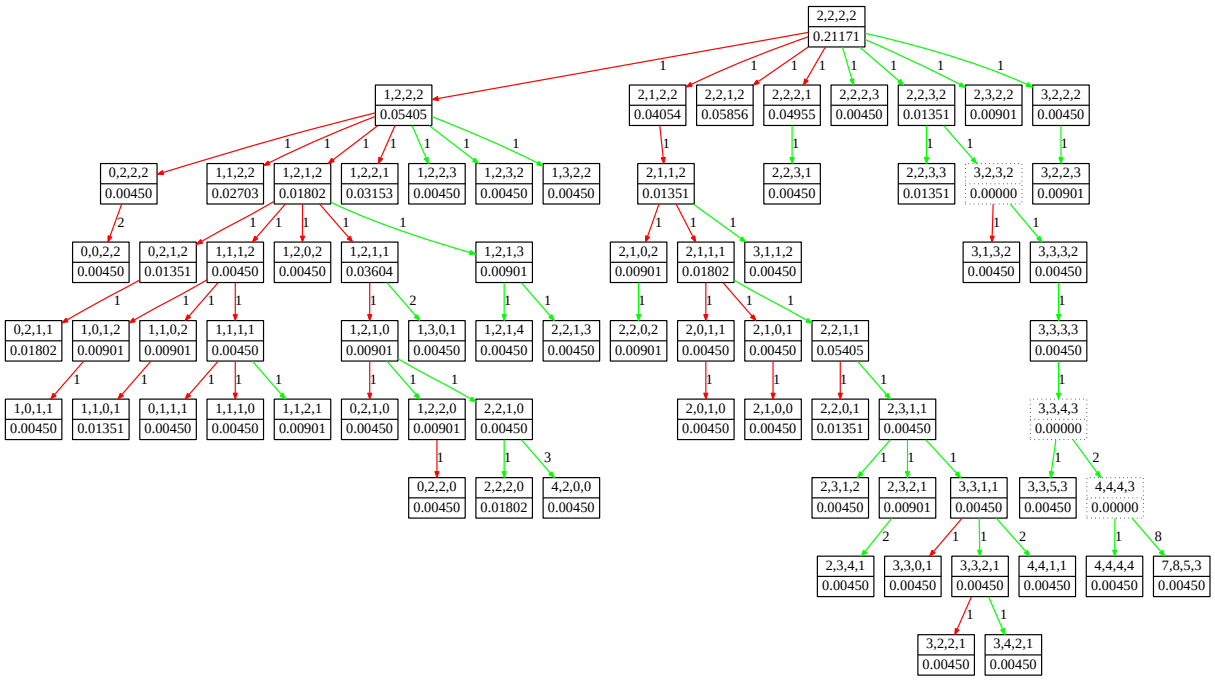

Figure SF19: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 7. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

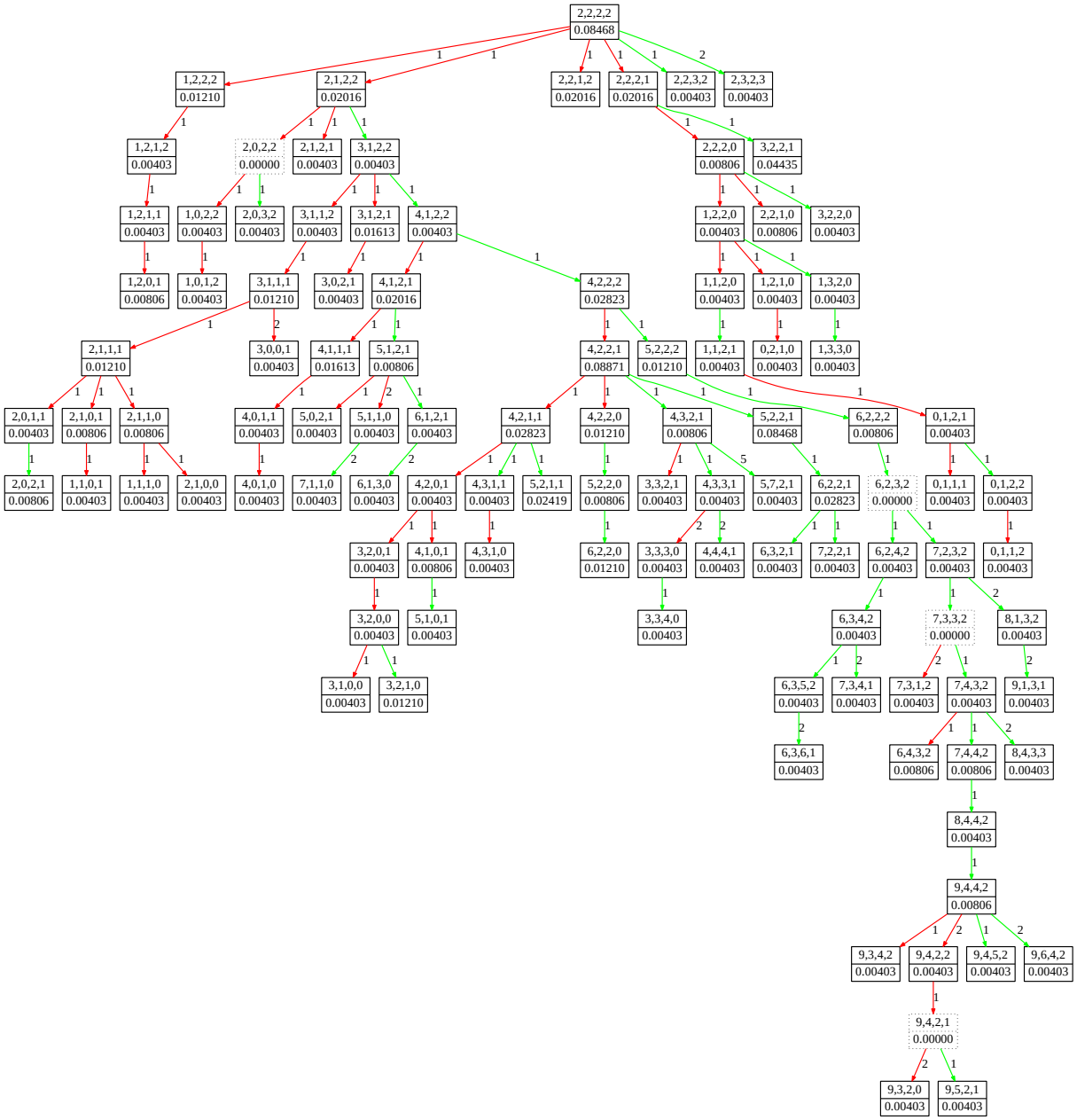

Figure SF20: Phylogenetic trees showing progression of primary stage cervical cancer in patient 8. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

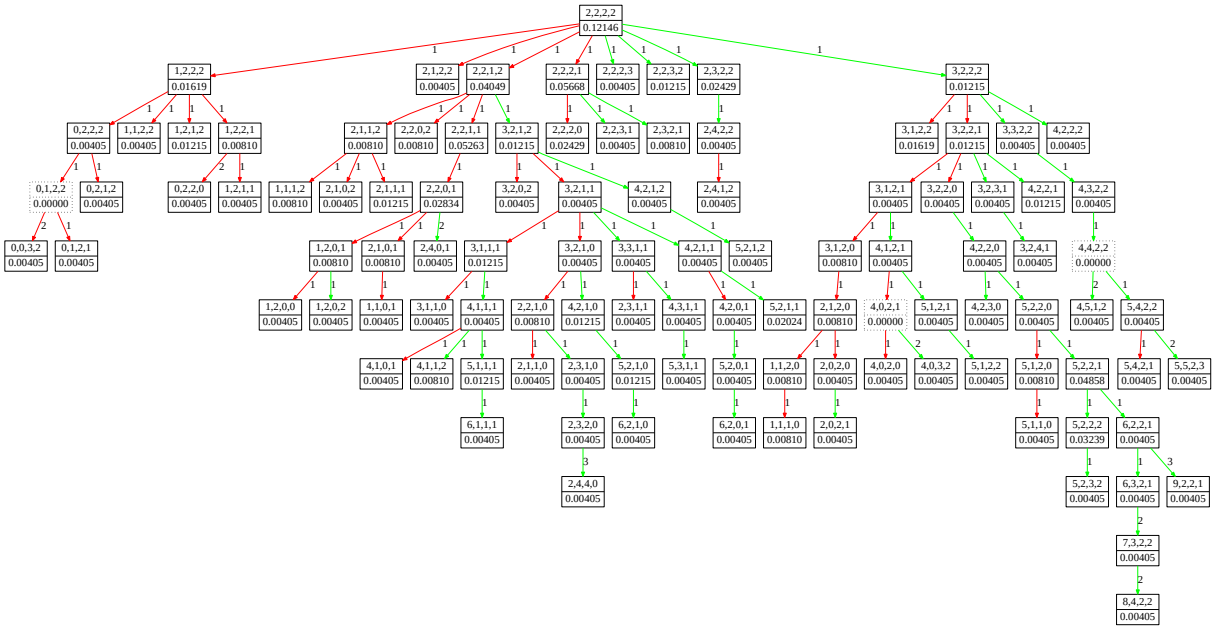

Figure SF21: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 8. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

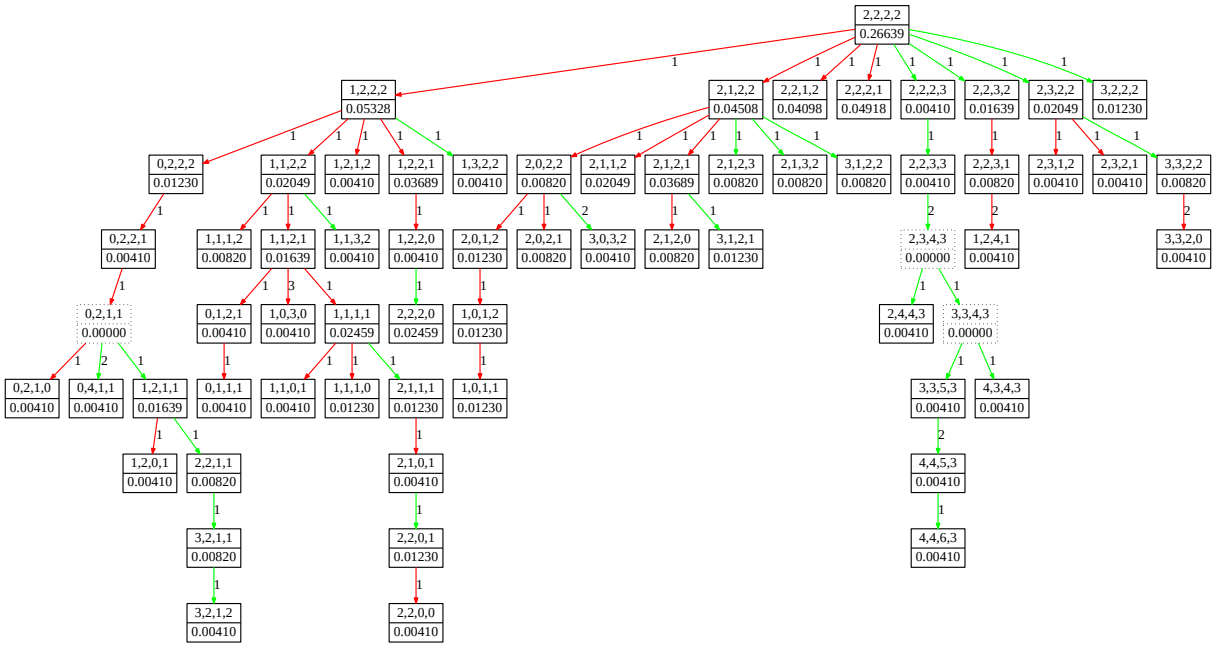

Figure SF22: Phylogenetic trees showing progression of primary stage cervical cancer in patient 9. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

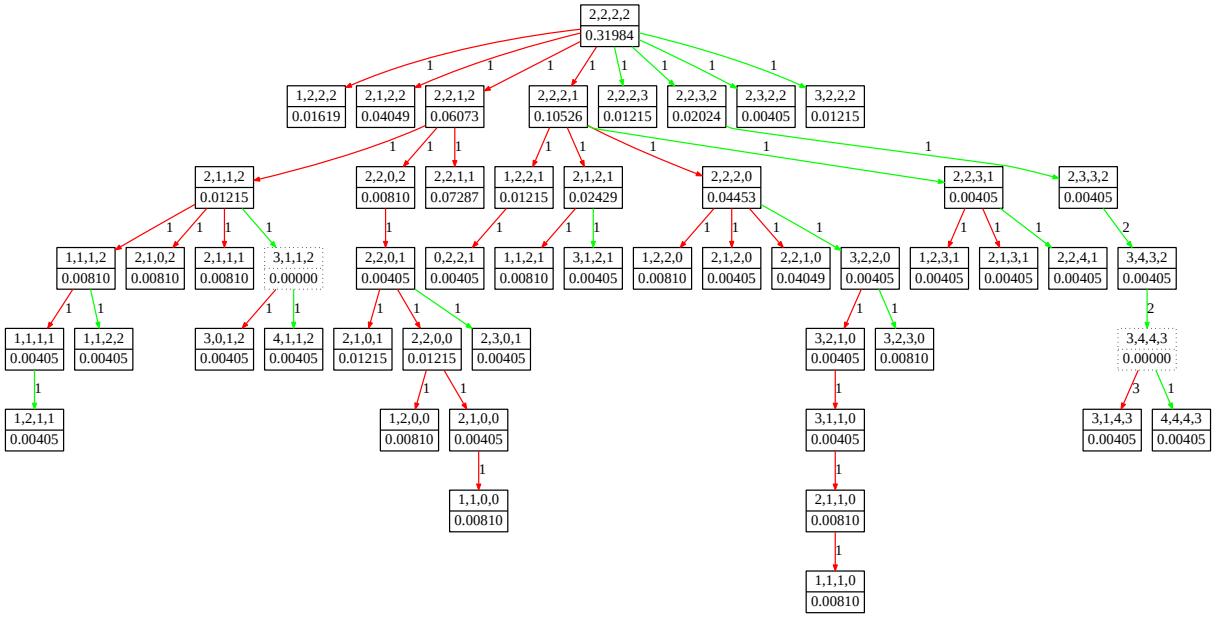

Figure SF23: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 9. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

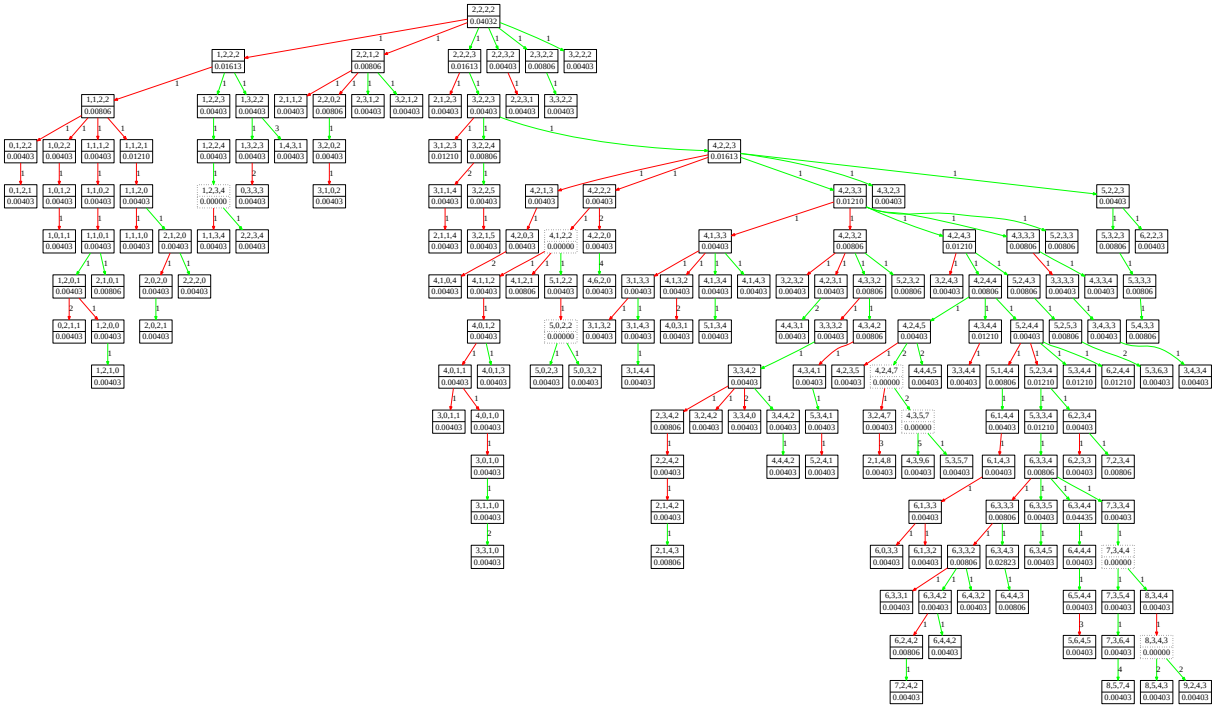

Figure SF24: Phylogenetic trees showing progression of primary stage cervical cancer in patient 10. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

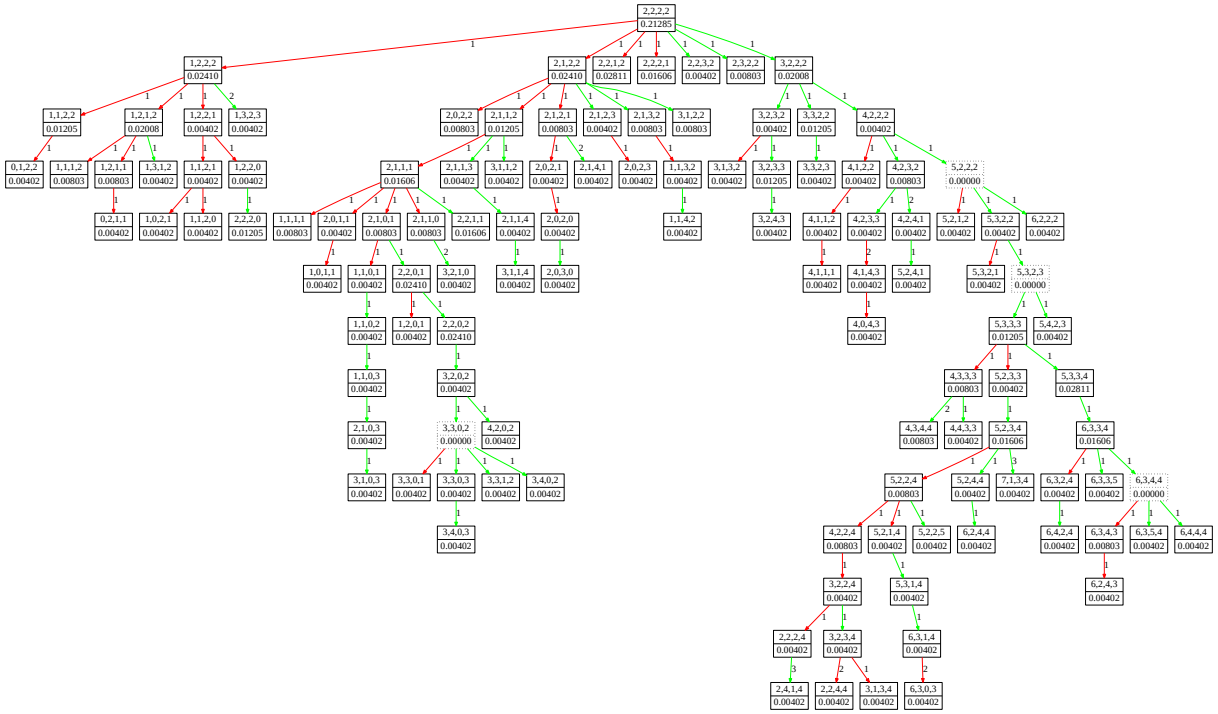

Figure SF25: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 10. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

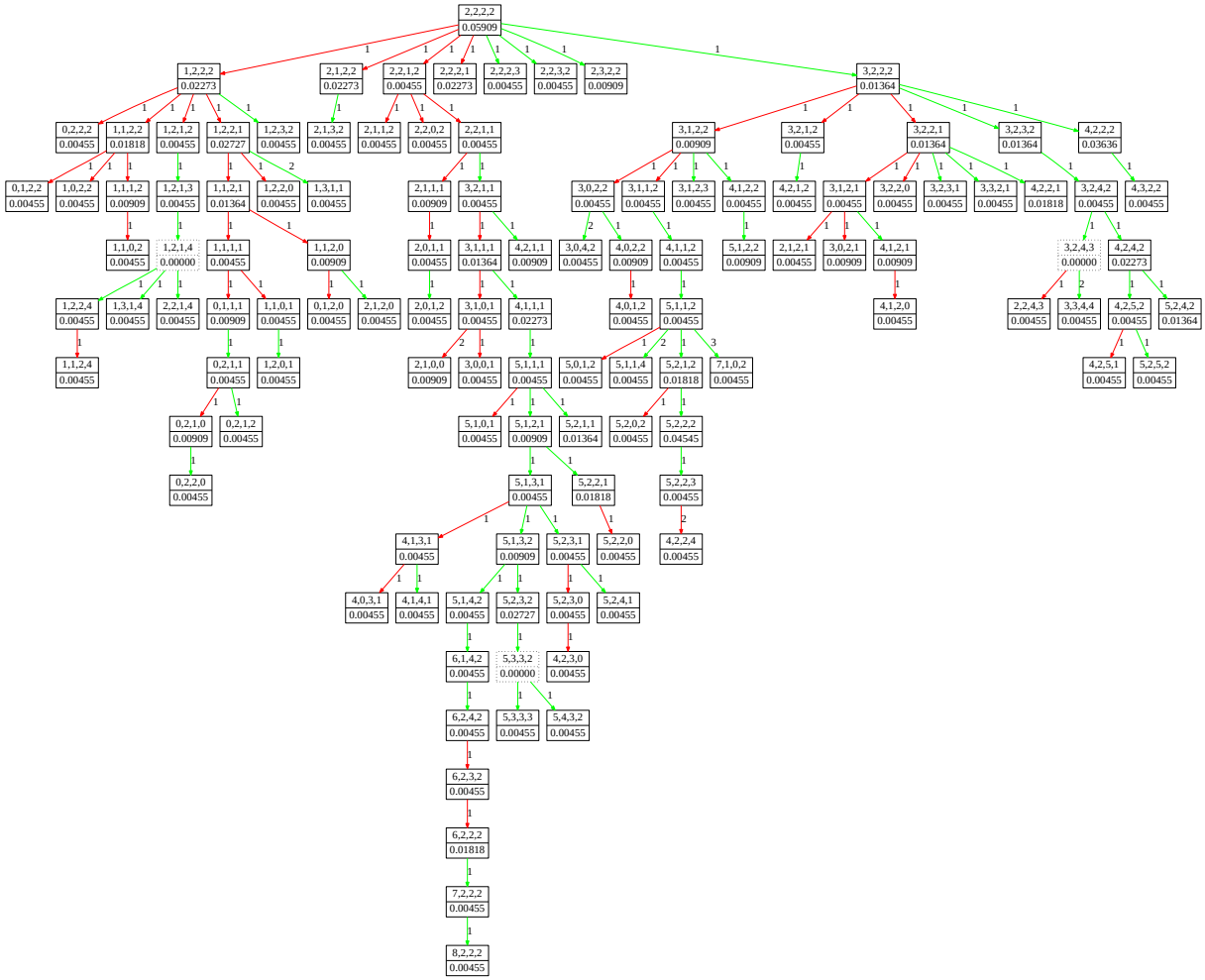

Figure SF26: Phylogenetic trees showing progression of primary stage cervical cancer in patient 11. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

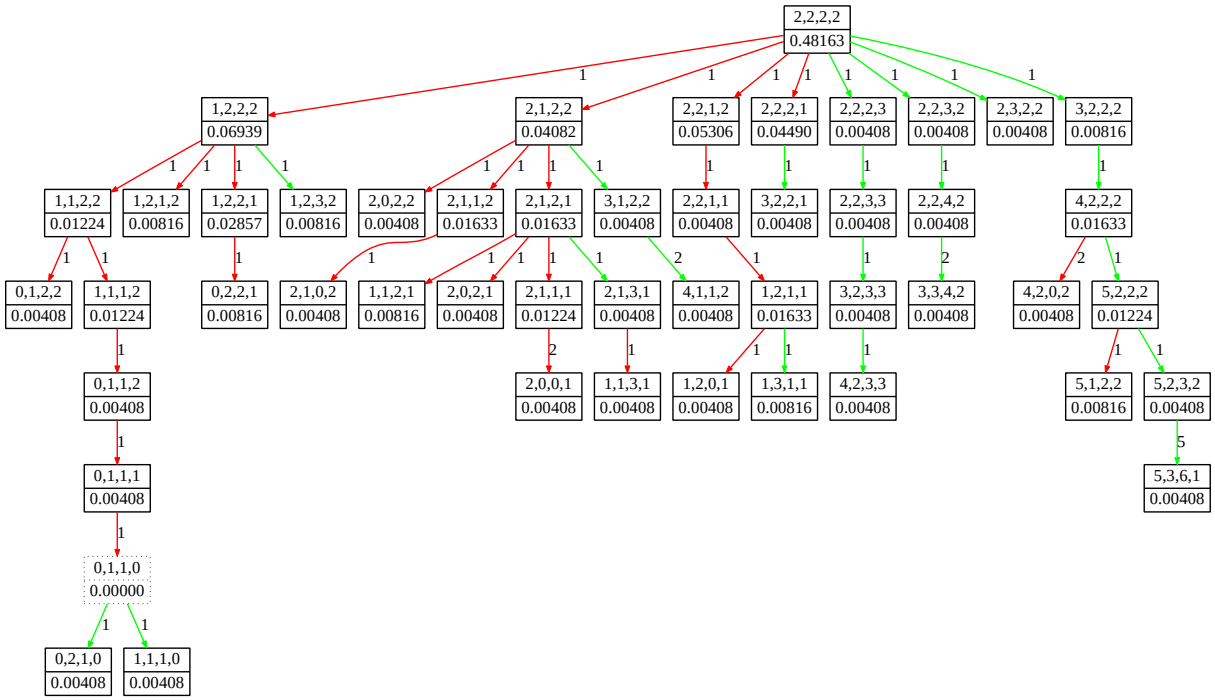

Figure SF27: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 11. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

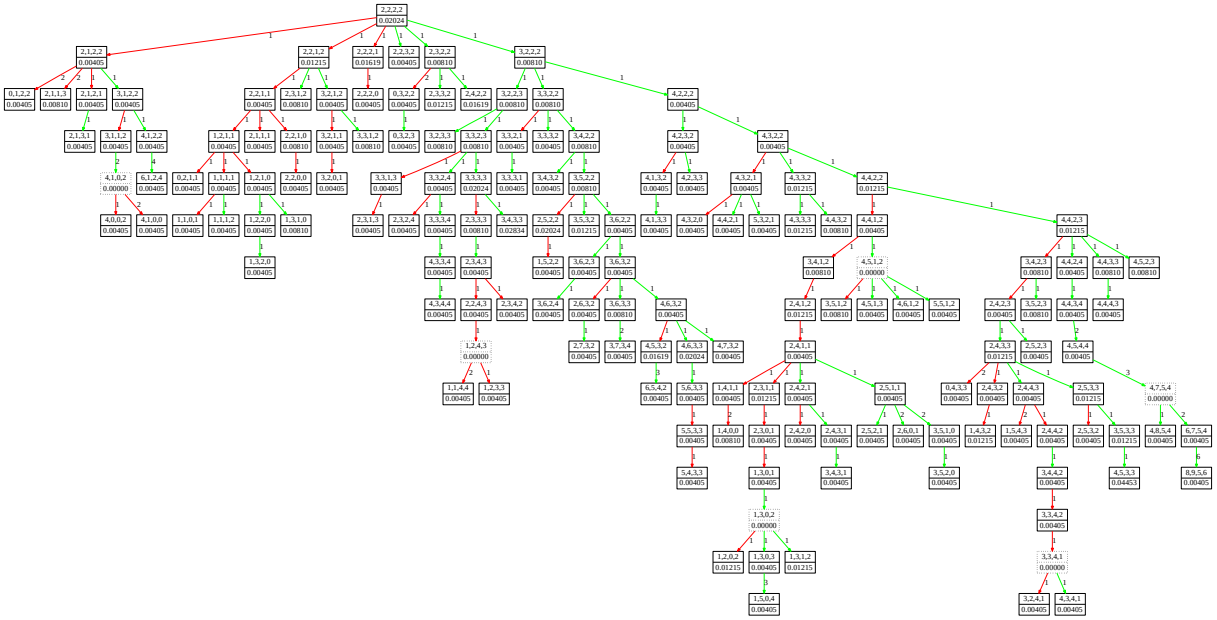

Figure SF28: Phylogenetic trees showing progression of primary stage cervical cancer in patient 12. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

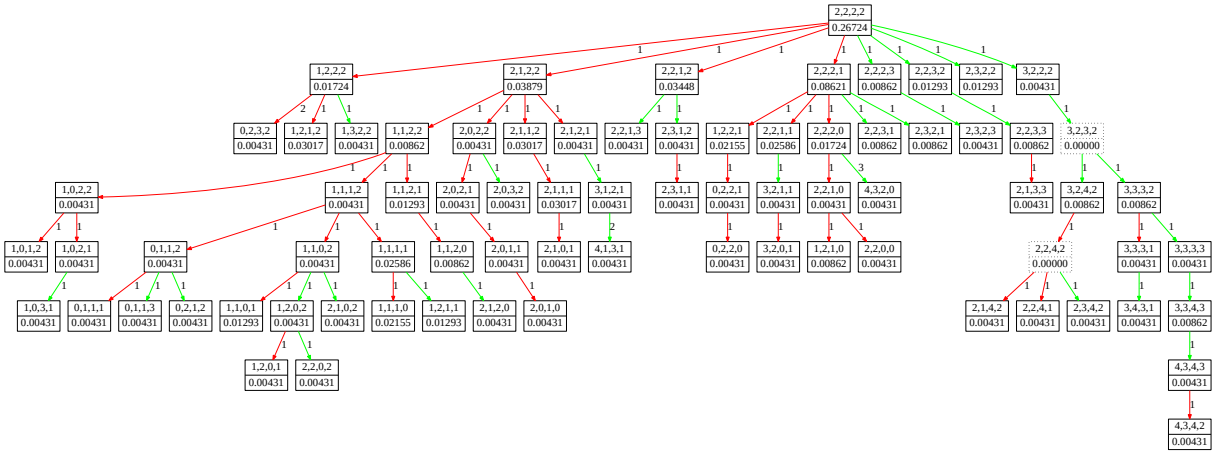

Figure SF29: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 12. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

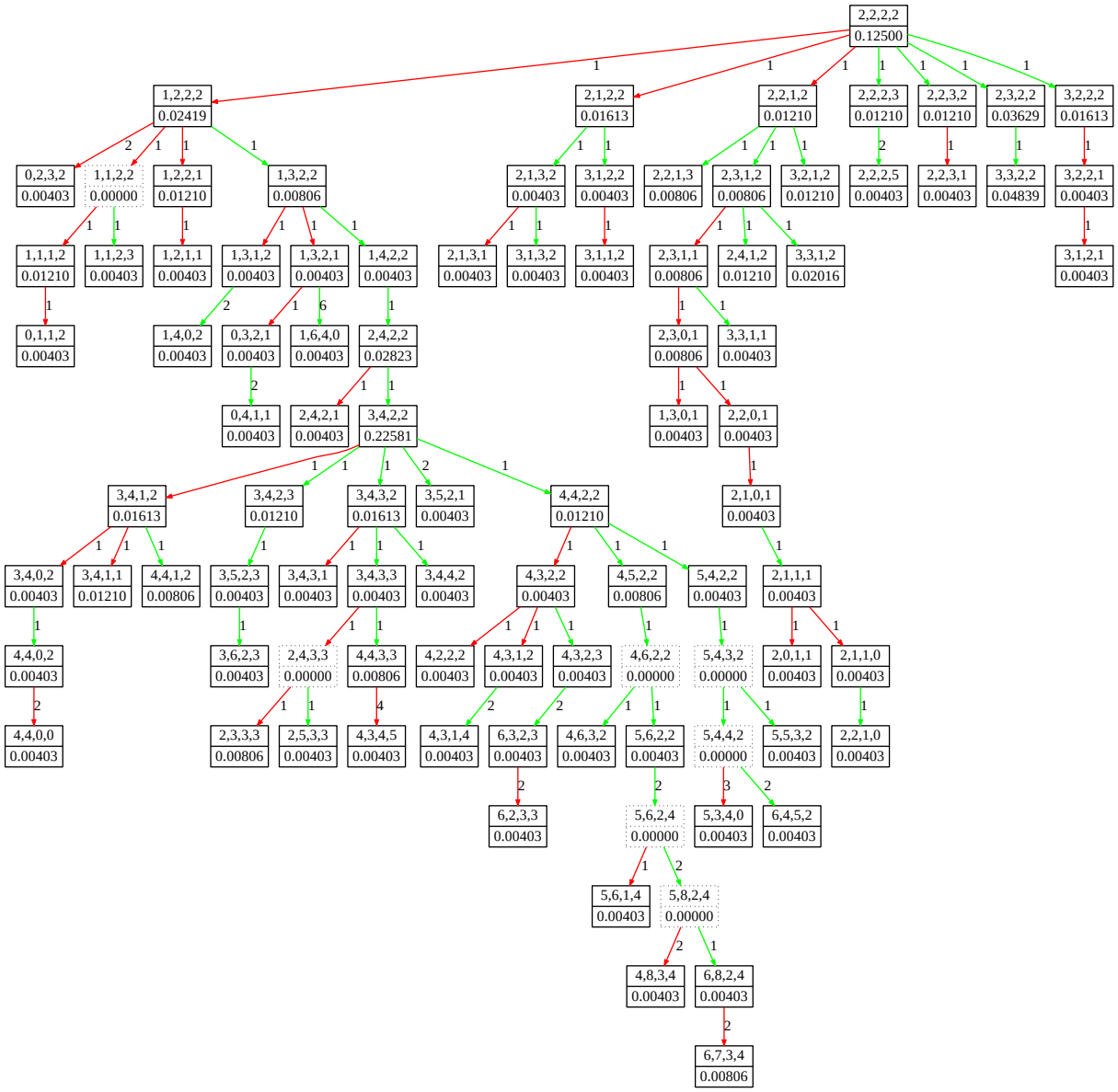

Figure SF30: Phylogenetic trees showing progression of primary stage cervical cancer in patient 13. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

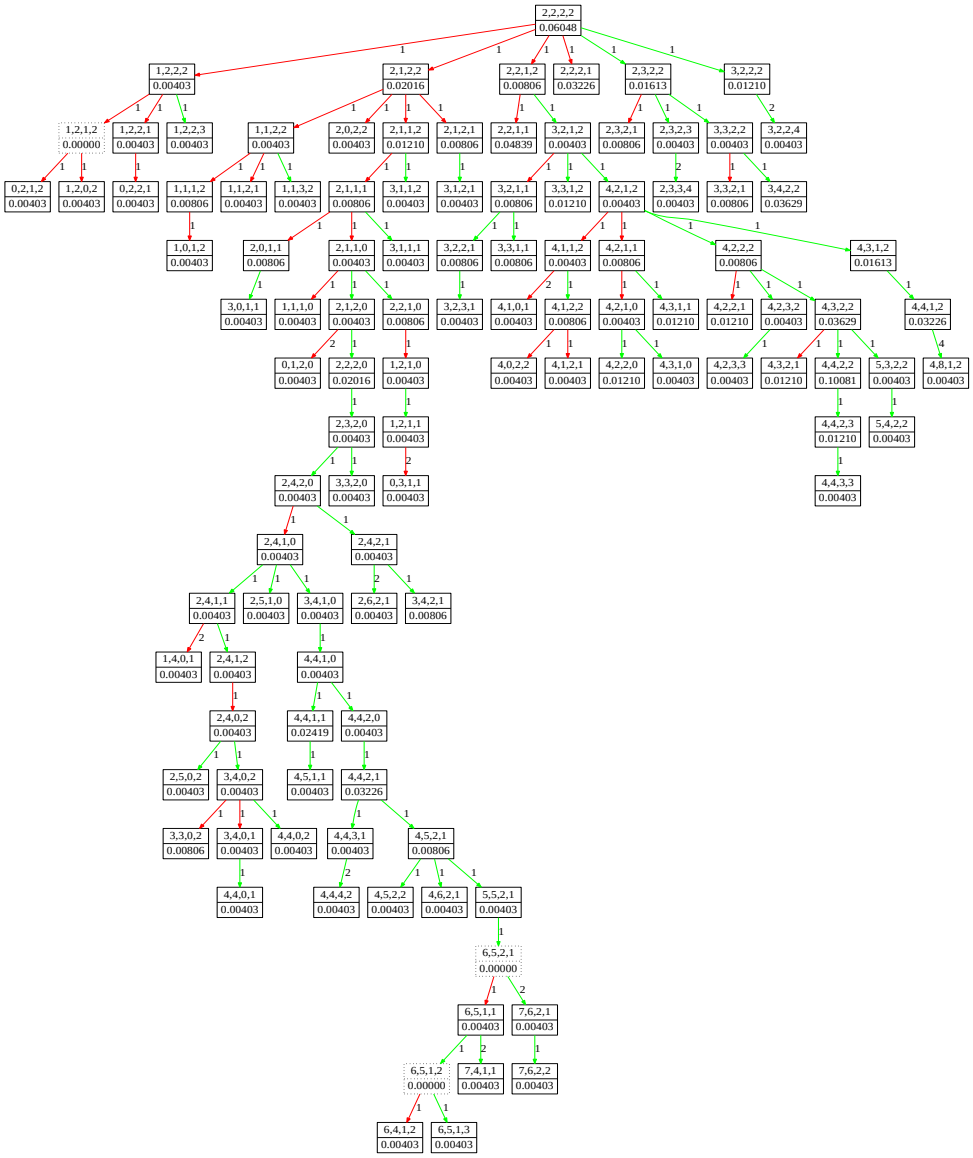

Figure SF31: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 13. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

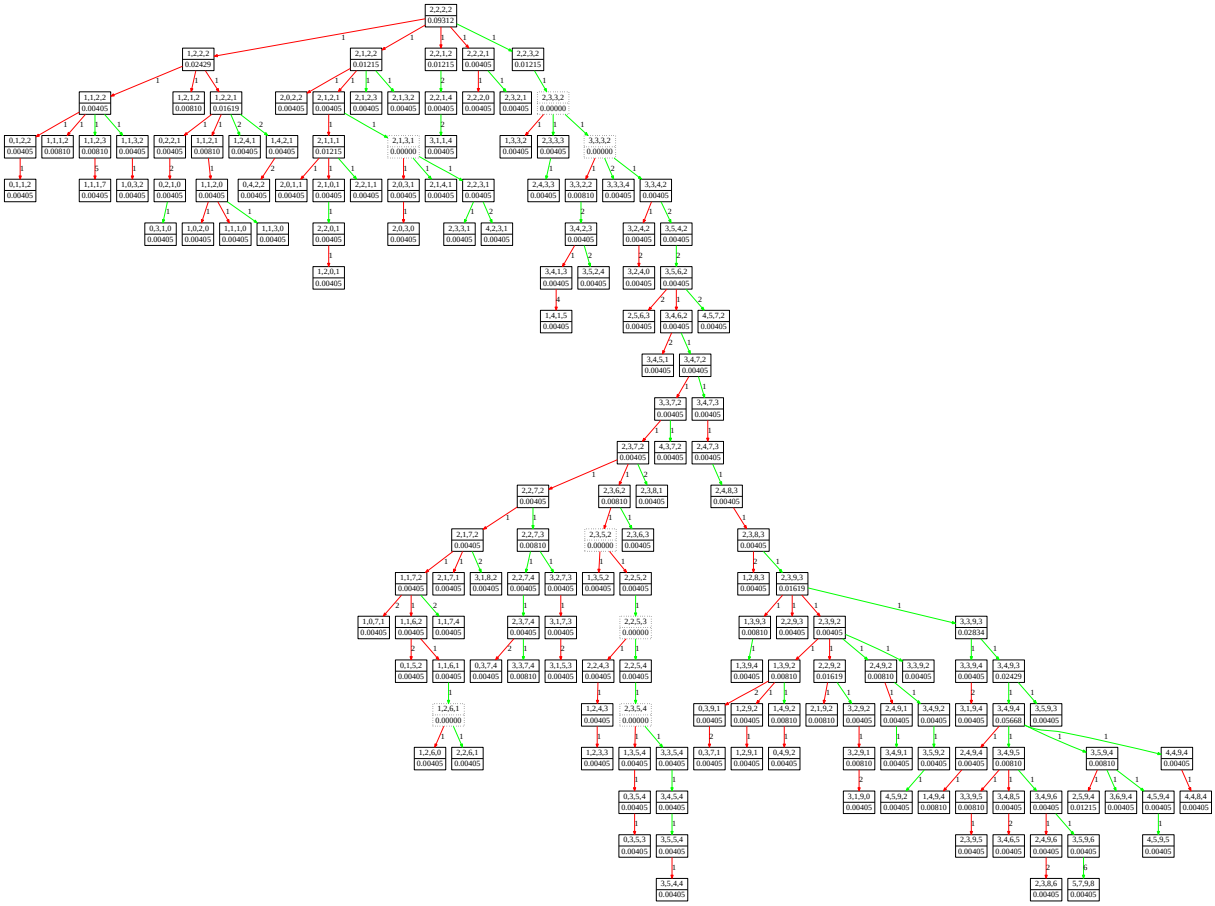

Figure SF32: Phylogenetic trees showing progression of primary stage cervical cancer in patient 14. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

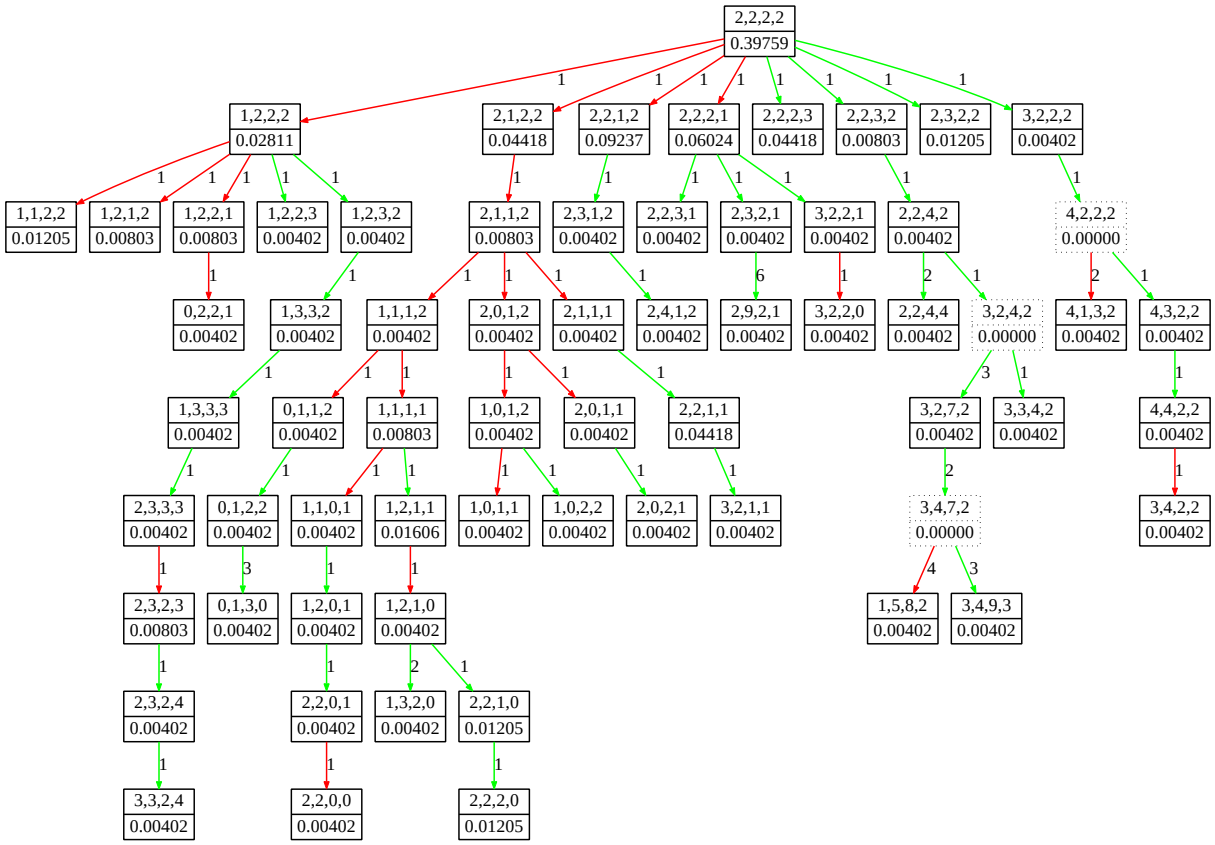

Figure SF33: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 14. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

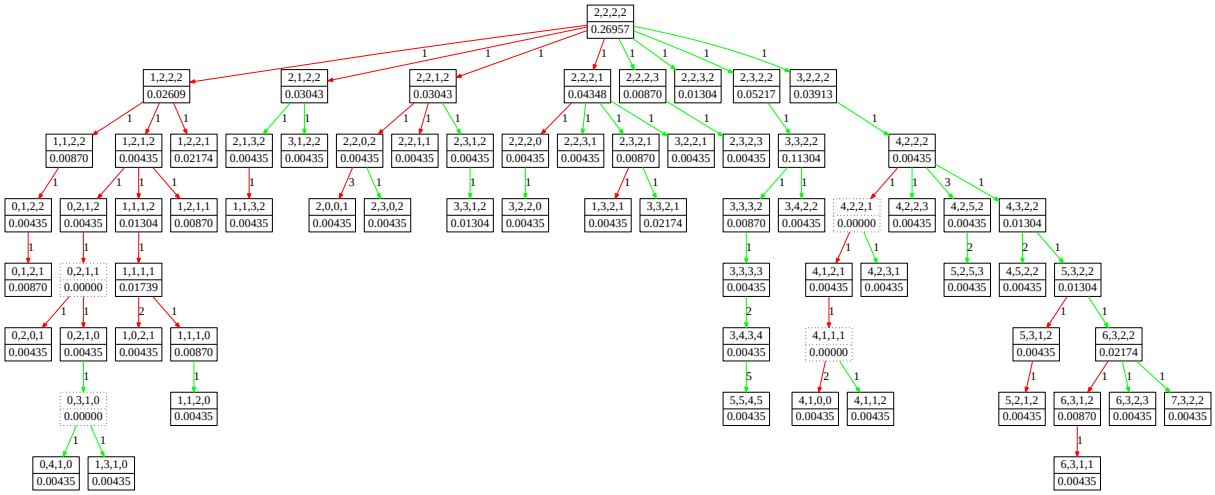

Figure SF34: Phylogenetic trees showing progression of primary stage cervical cancer in patient 15. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

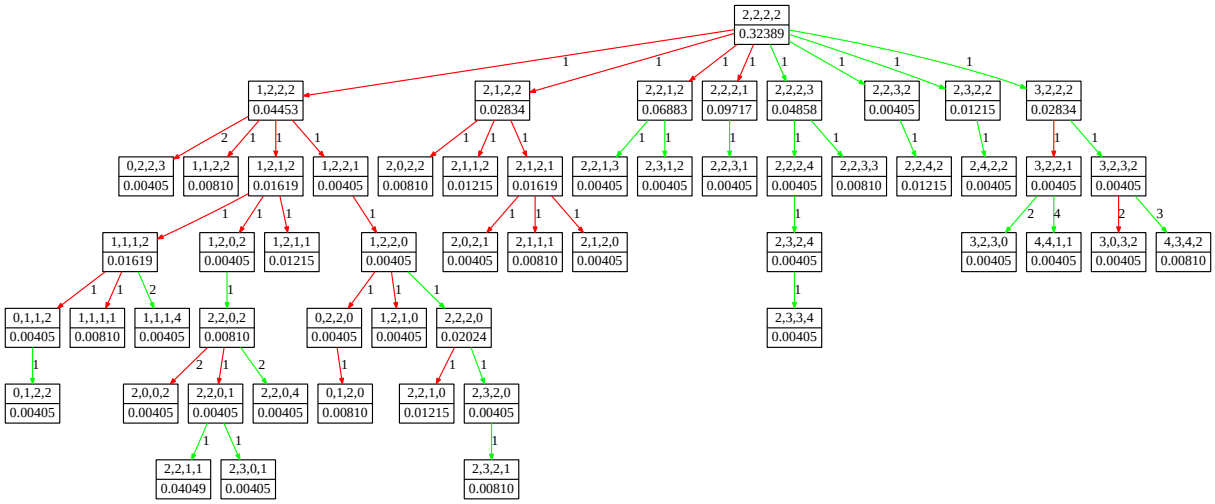

Figure SF35: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 15. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

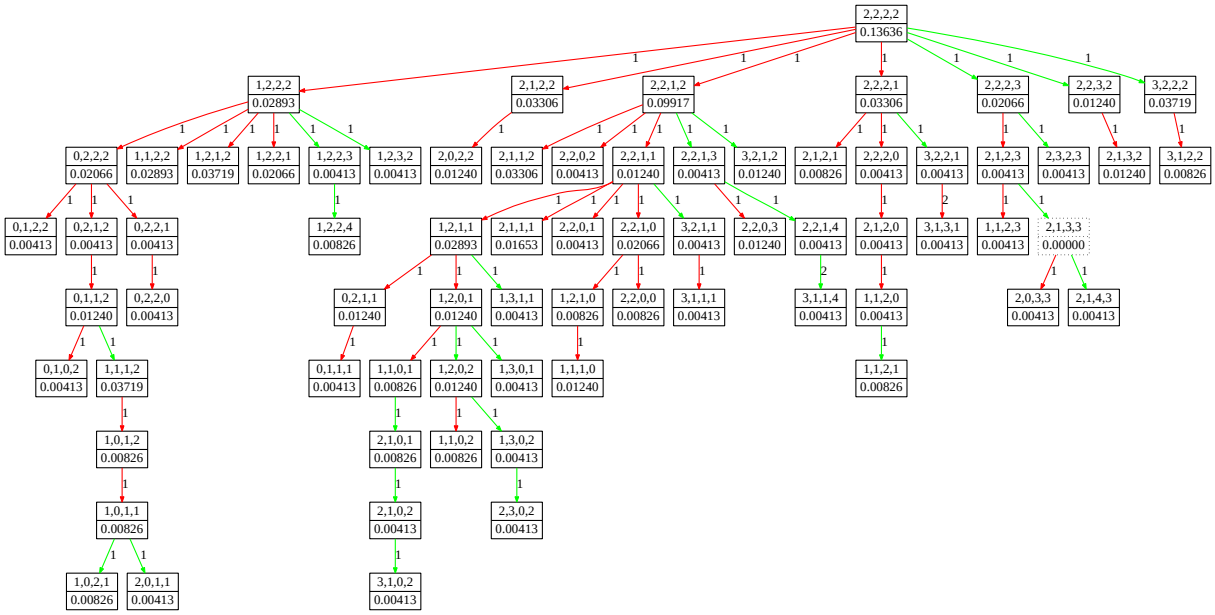

Figure SF36: Phylogenetic trees showing progression of primary stage cervical cancer in patient 16. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

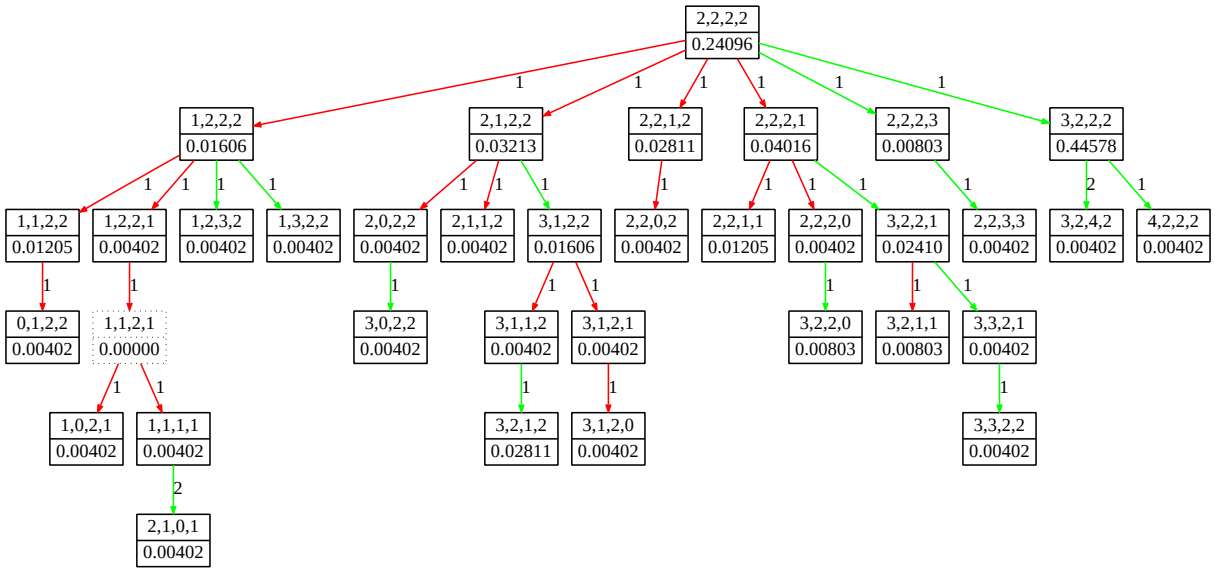

Figure SF37: Phylogenetic trees showing progression of metastasis stage cervical cancer in patient 16. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

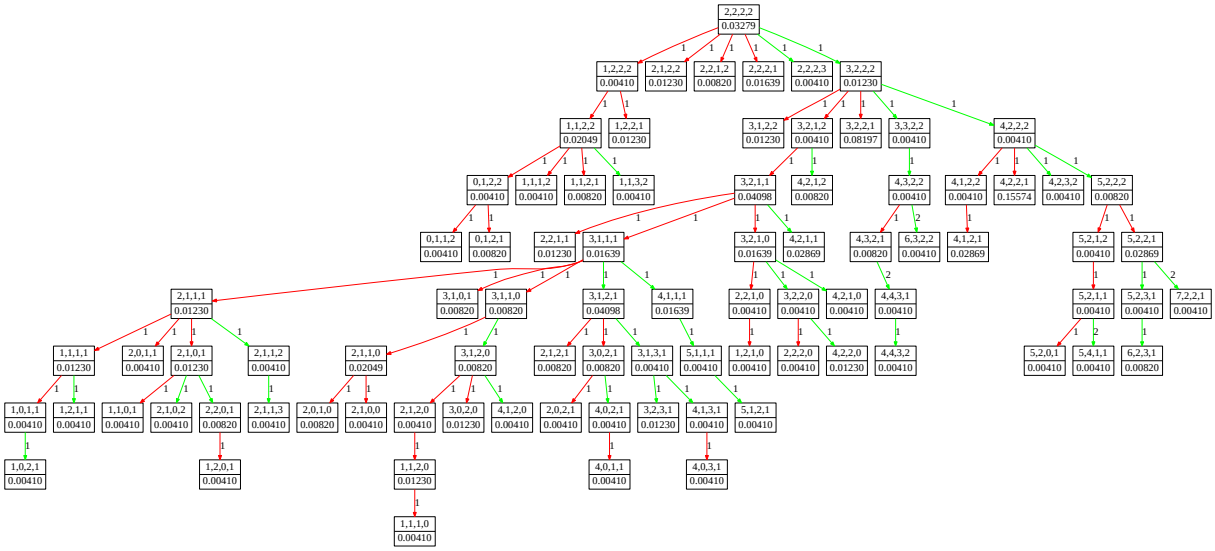

Figure SF38: Phylogenetic trees showing progression of primary stage cervical cancer in patient 17. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

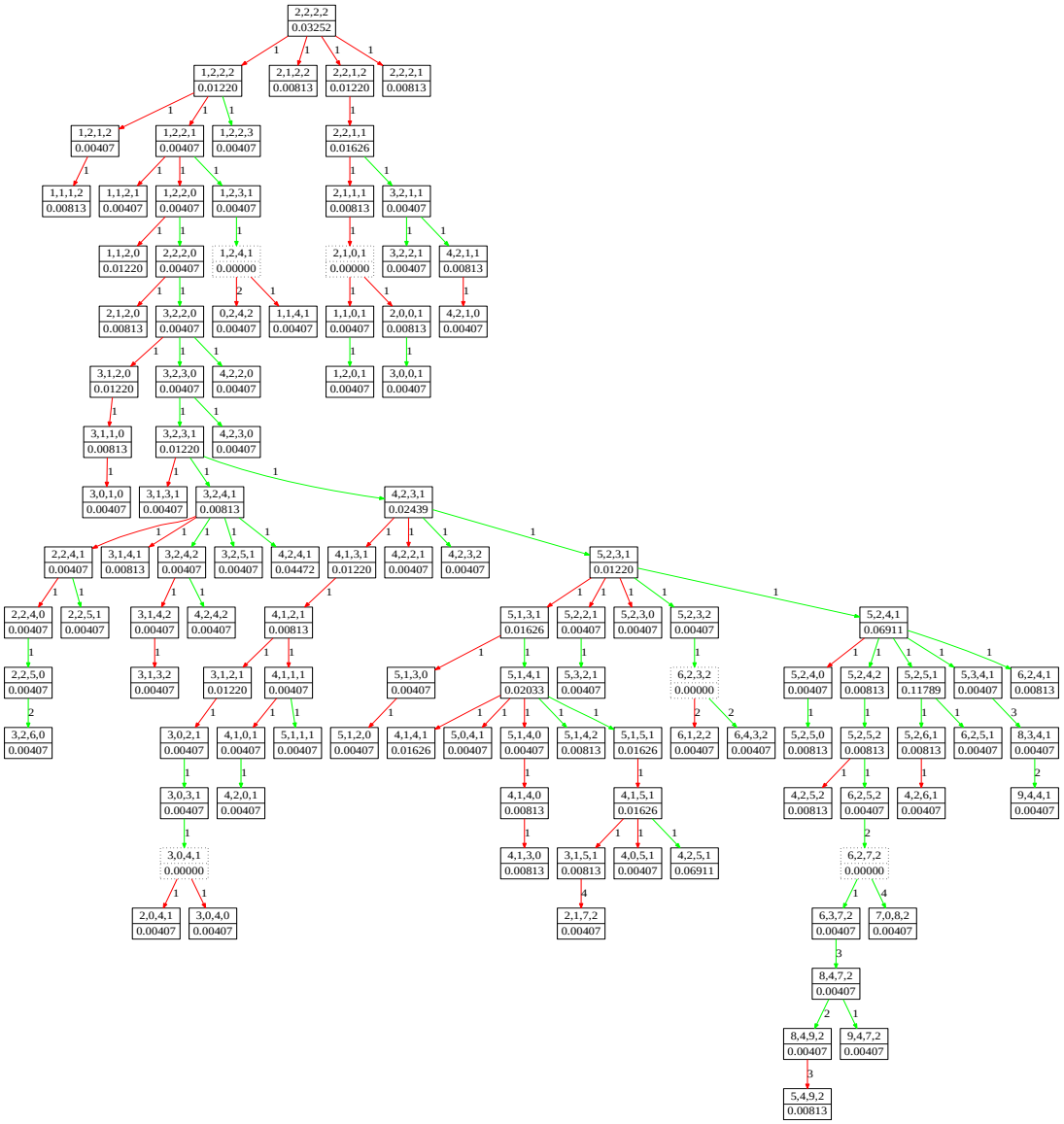

Figure SF39: Phylogenetic trees showing progression of primary stage cervical cancer in patient 18. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

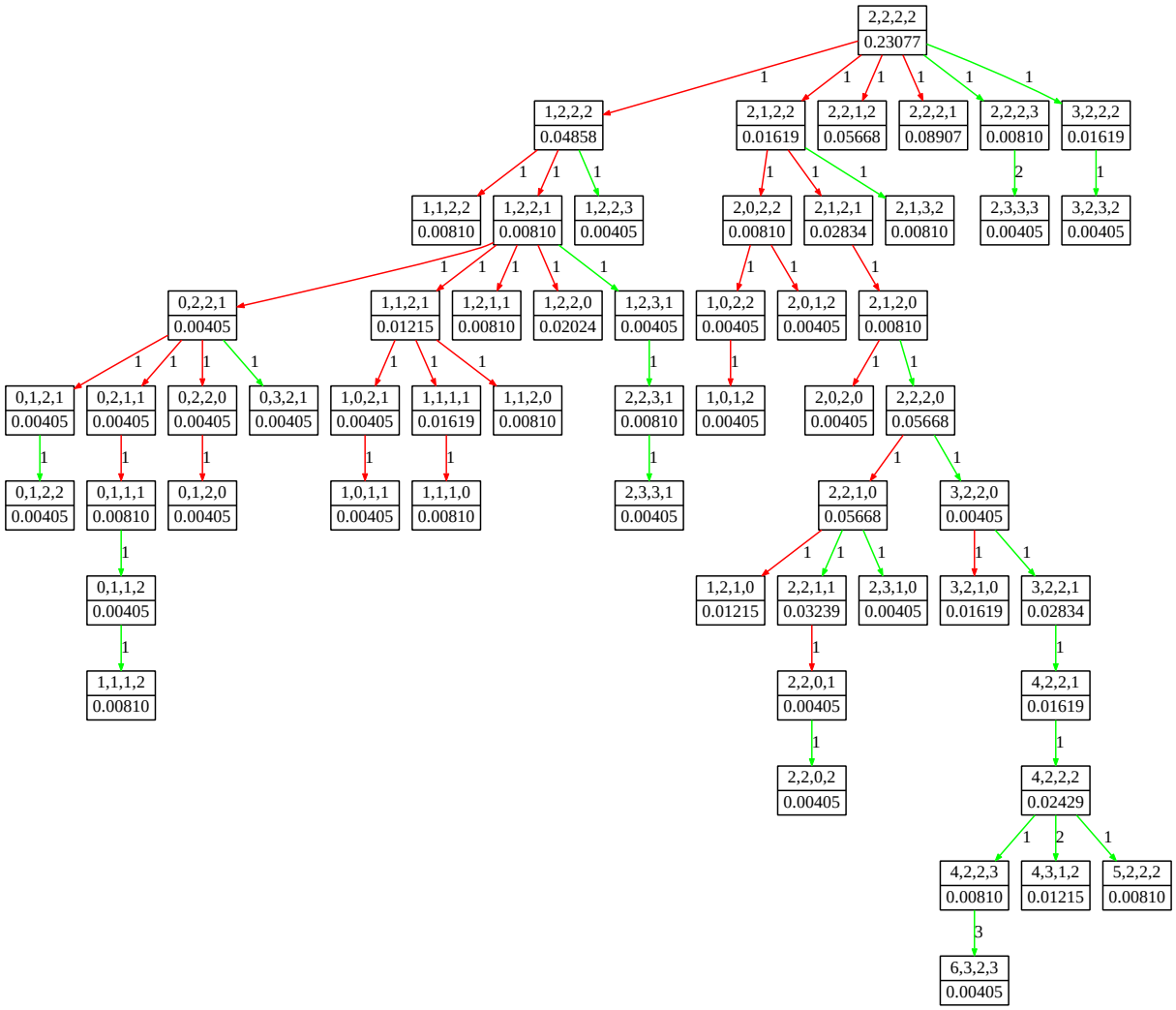

Figure SF40: Phylogenetic trees showing progression of primary stage cervical cancer in patient 19. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

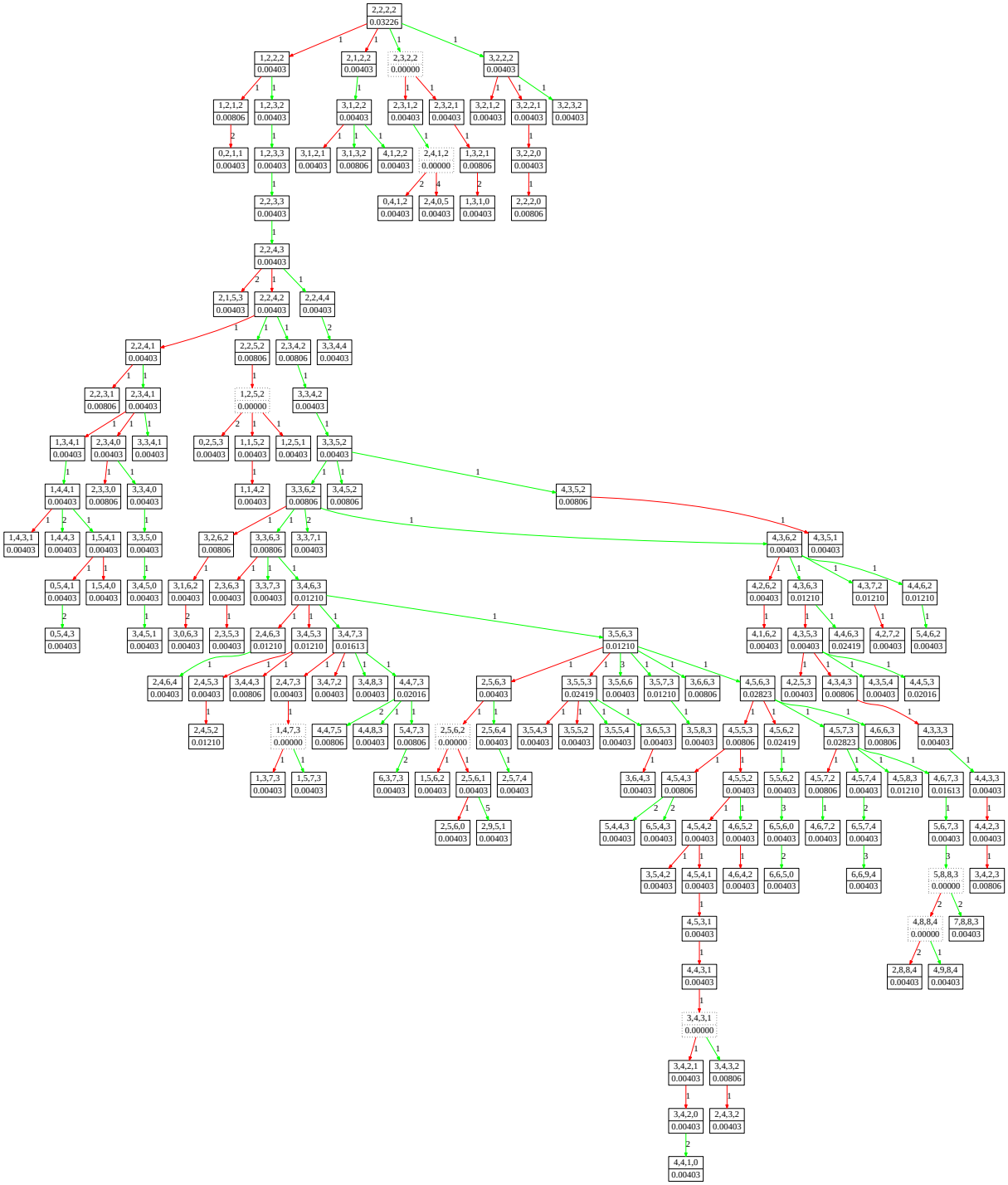

Figure SF41: Phylogenetic trees showing progression of primary stage cervical cancer in patient 20. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

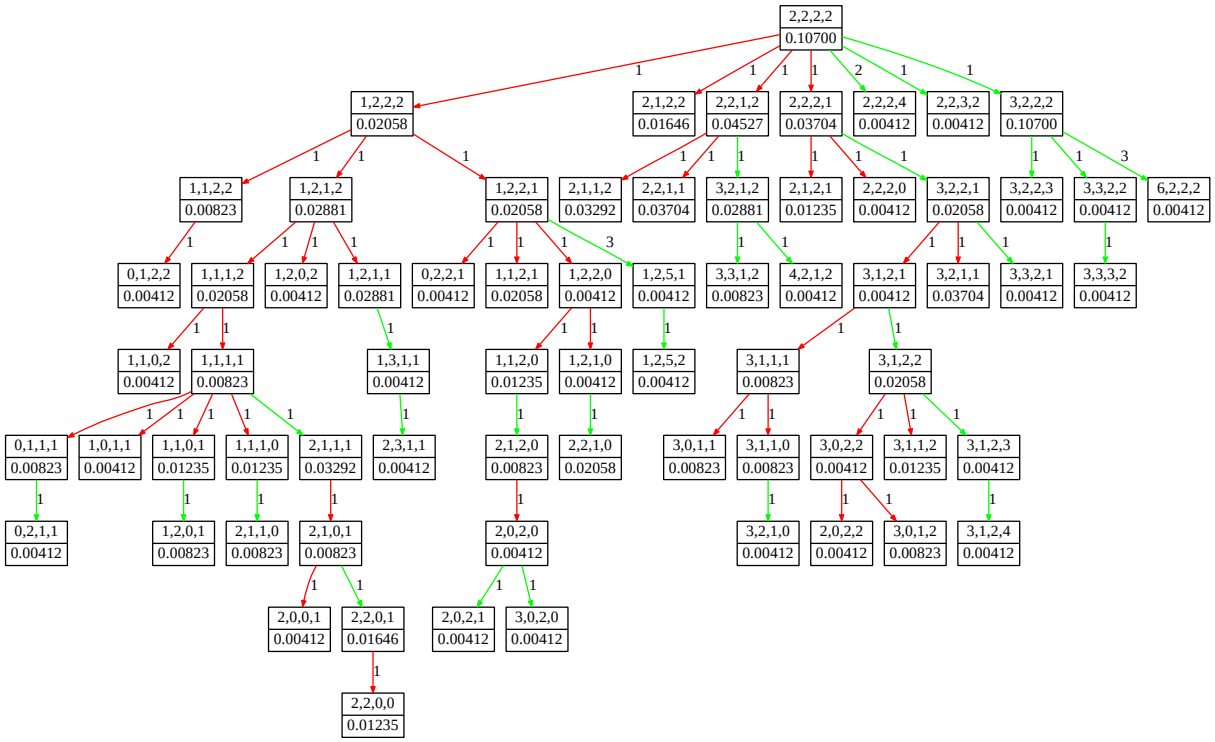

Figure SF42: Phylogenetic trees showing progression of primary stage cervical cancer in patient 21. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

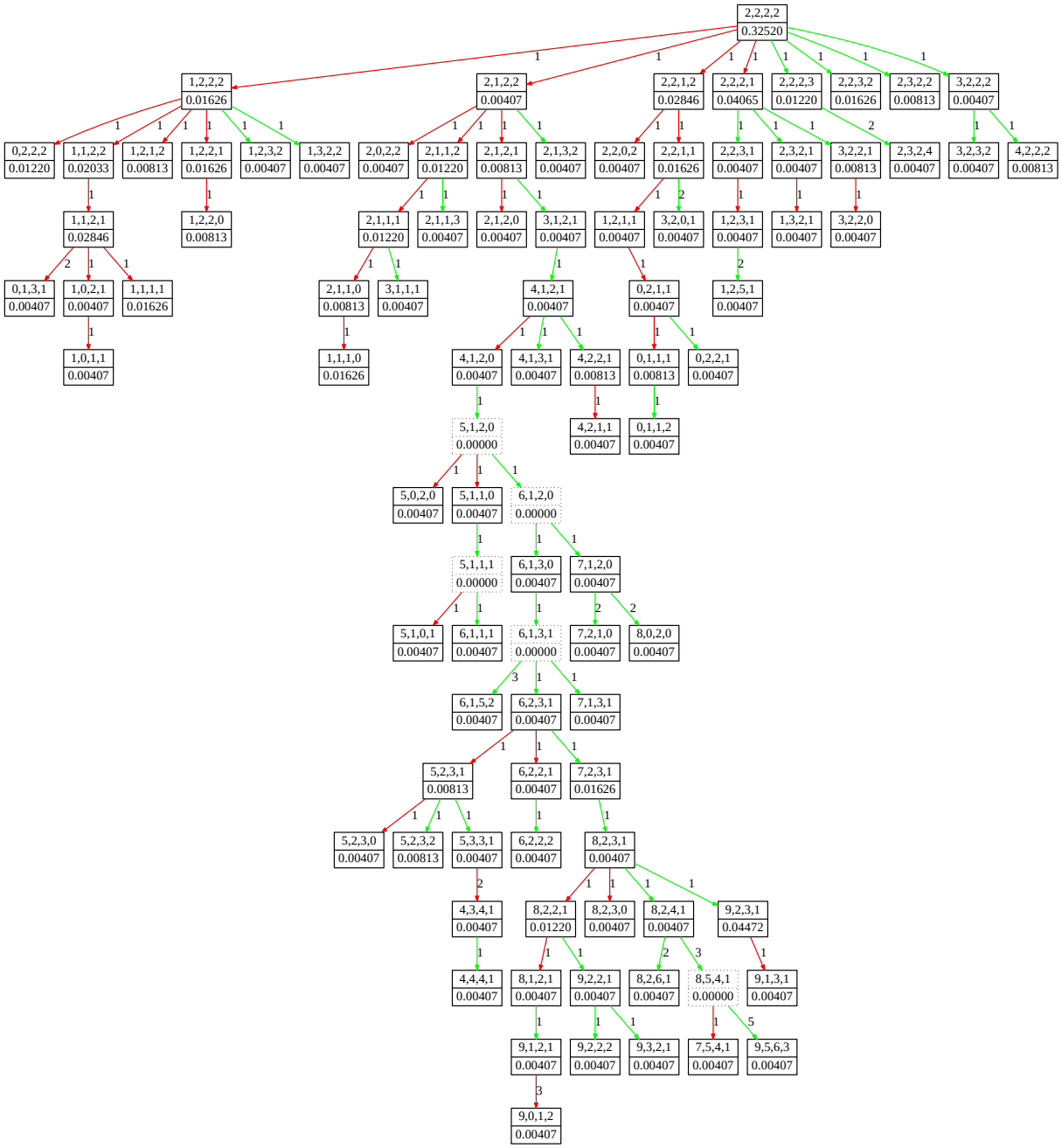

Figure SF43: Phylogenetic trees showing progression of primary stage cervical cancer in patient 22. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

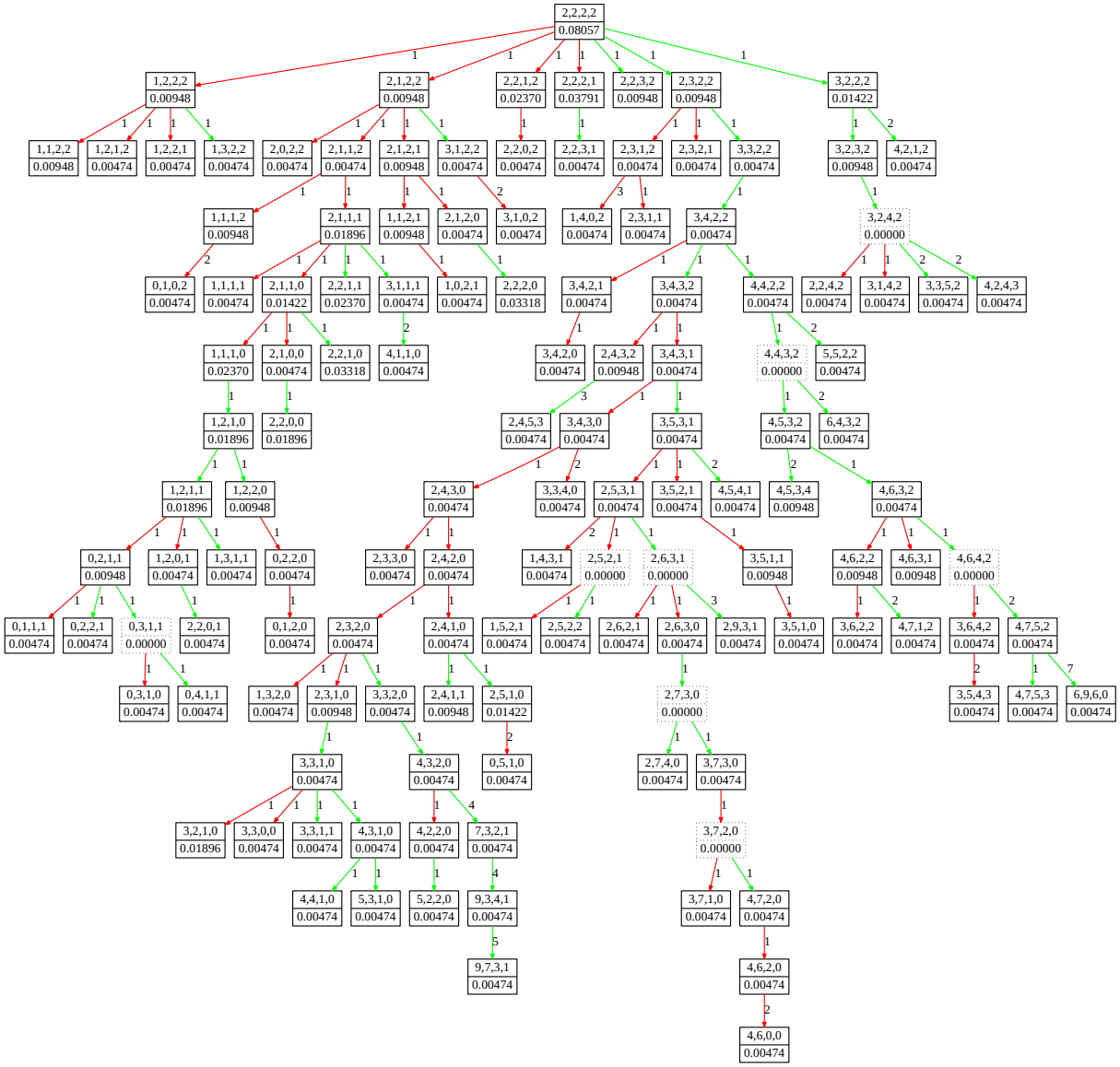

Figure SF44: Phylogenetic trees showing progression of primary stage cervical cancer in patient 23. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

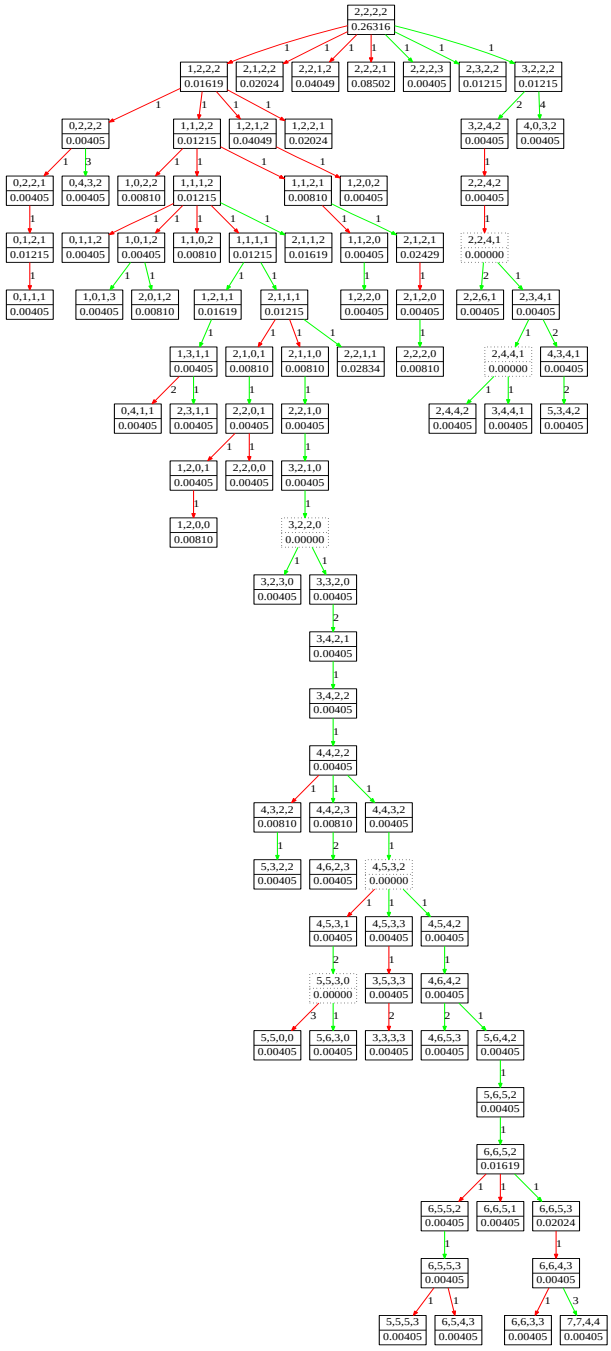

Figure SF45: Phylogenetic trees showing progression of primary stage cervical cancer in patient 24. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

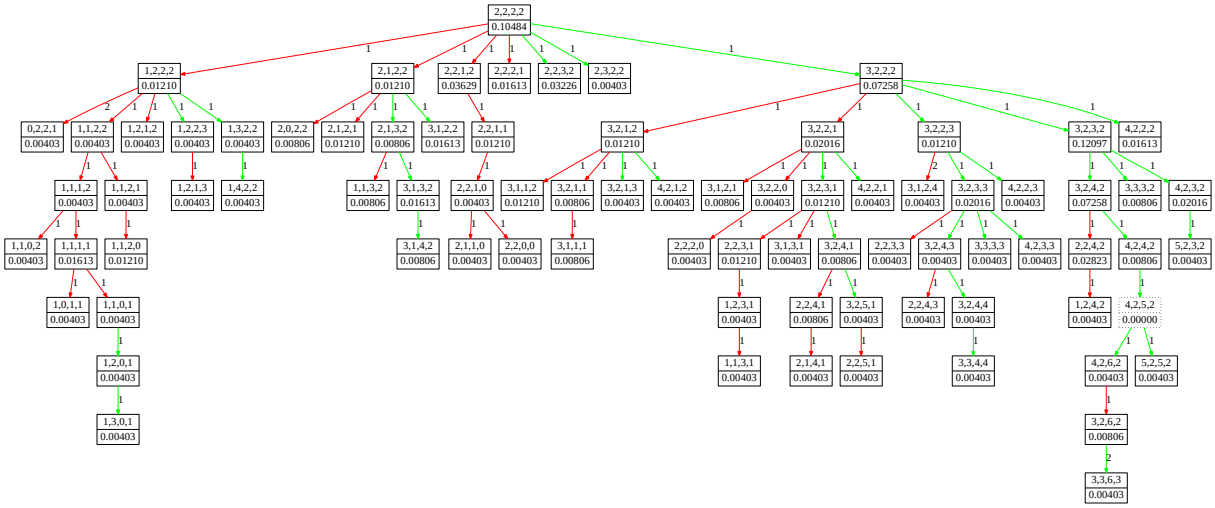

Figure SF46: Phylogenetic trees showing progression of primary stage cervical cancer in patient 25. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

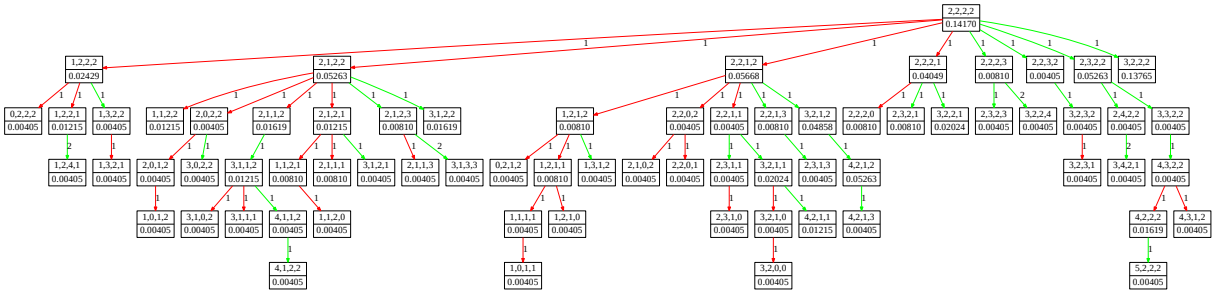

Figure SF47: Phylogenetic trees showing progression of primary stage cervical cancer in patient 26. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

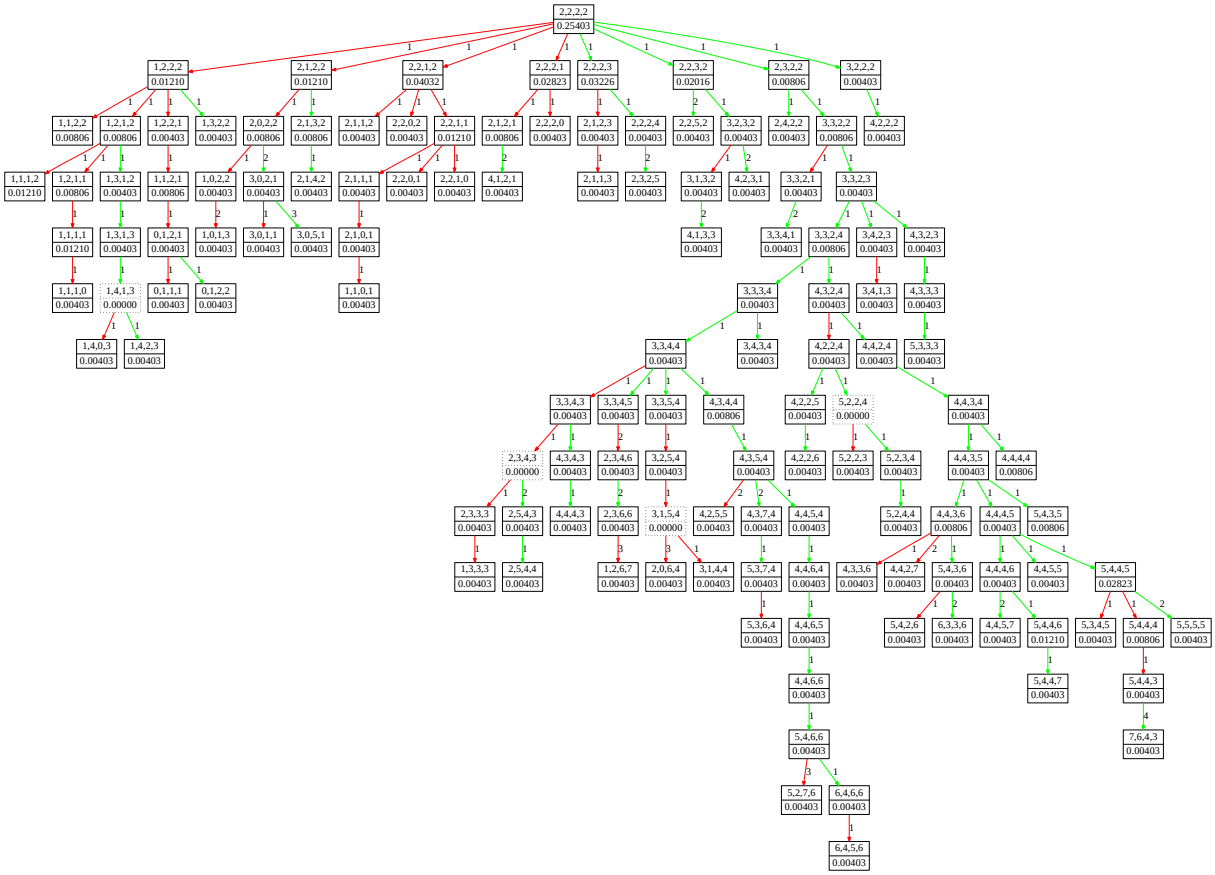

Figure SF48: Phylogenetic trees showing progression of primary stage cervical cancer in patient 27. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

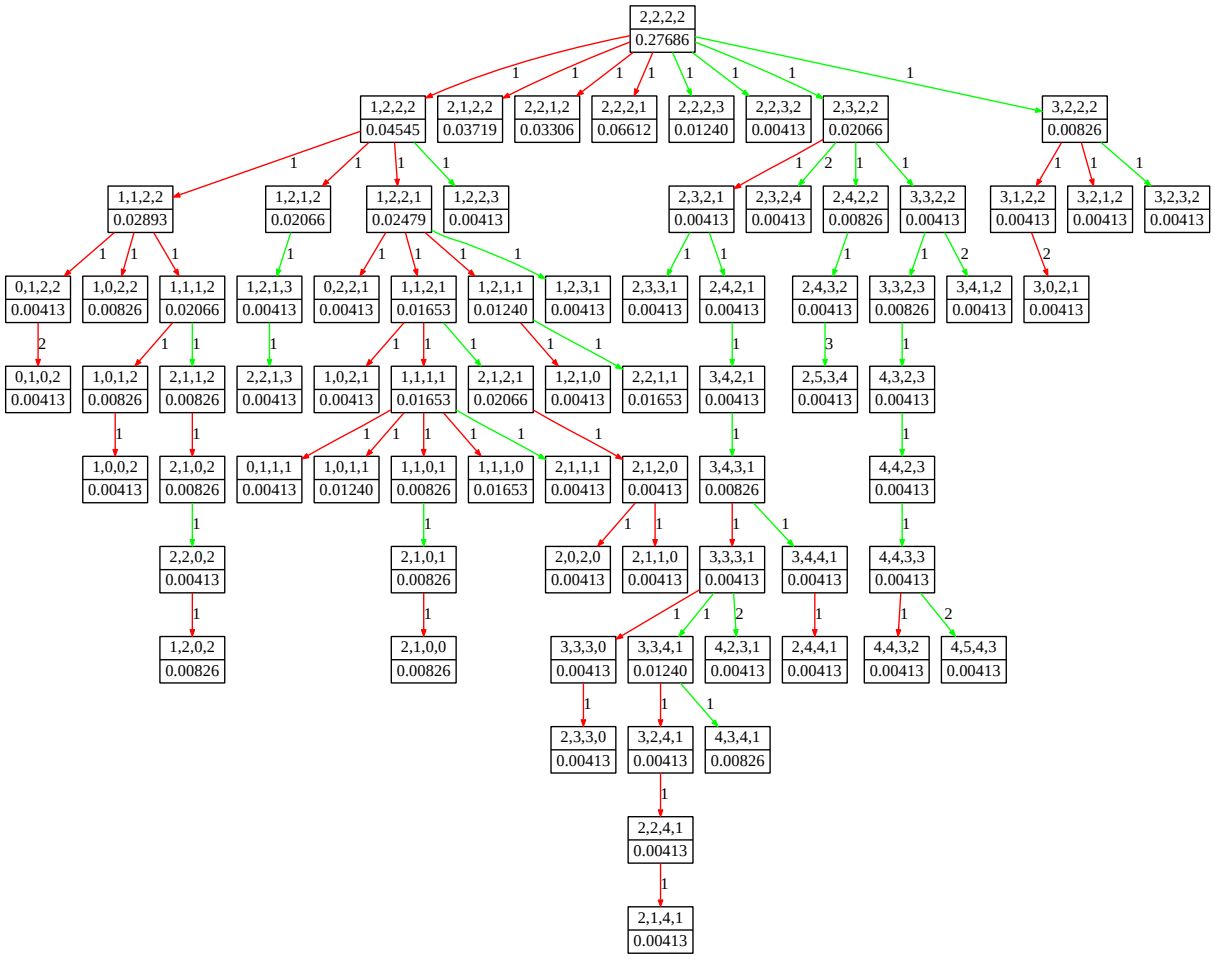

Figure SF49: Phylogenetic trees showing progression of primary stage cervical cancer in patient 28. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

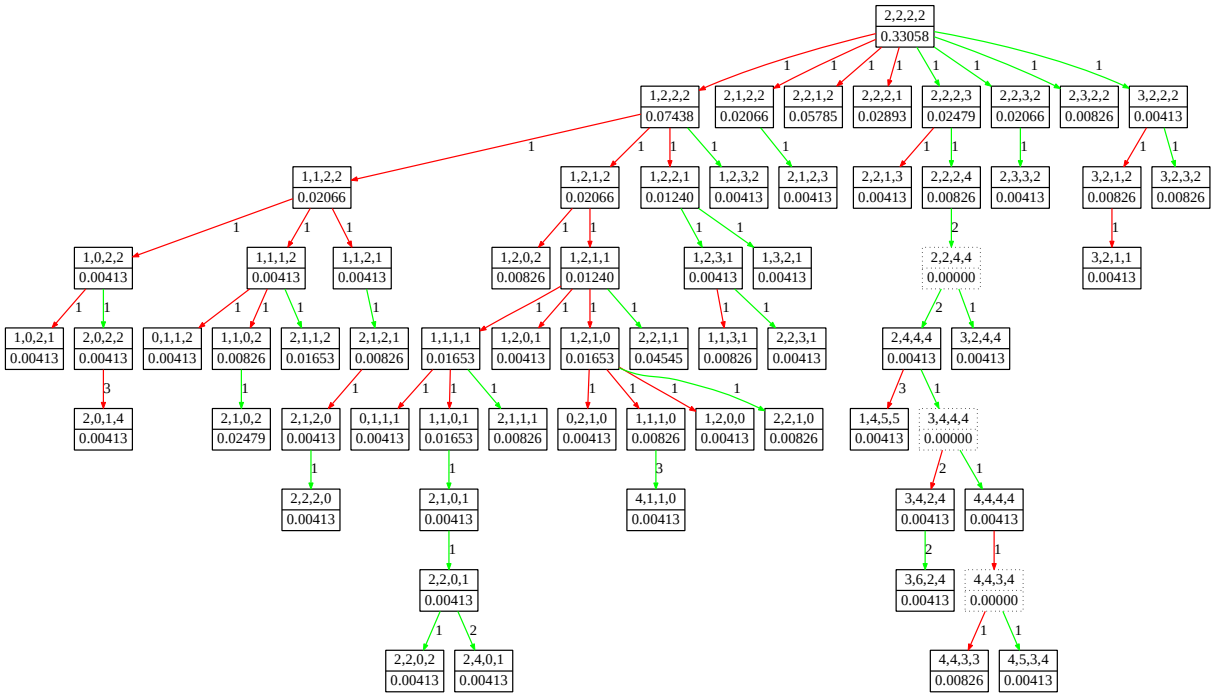

Figure SF50: Phylogenetic trees showing progression of primary stage cervical cancer in patient 29. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

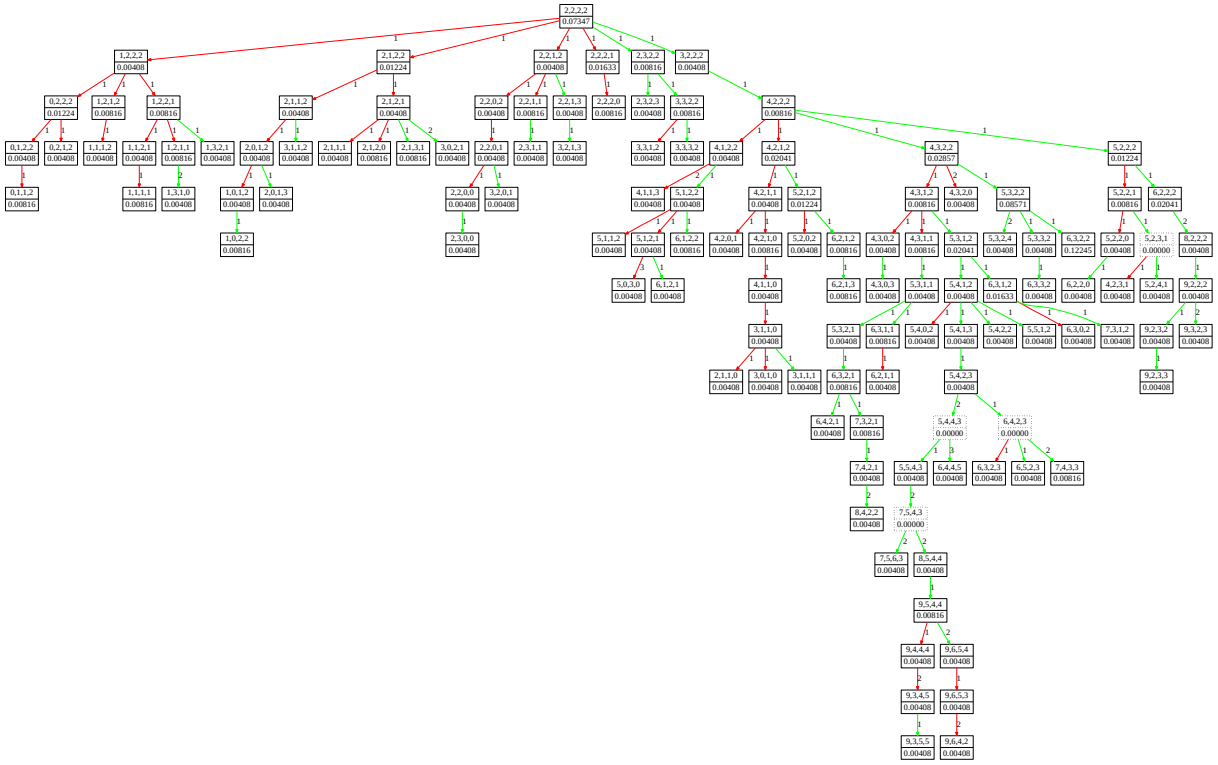

Figure SF51: Phylogenetic trees showing progression of primary stage cervical cancer in patient 30. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

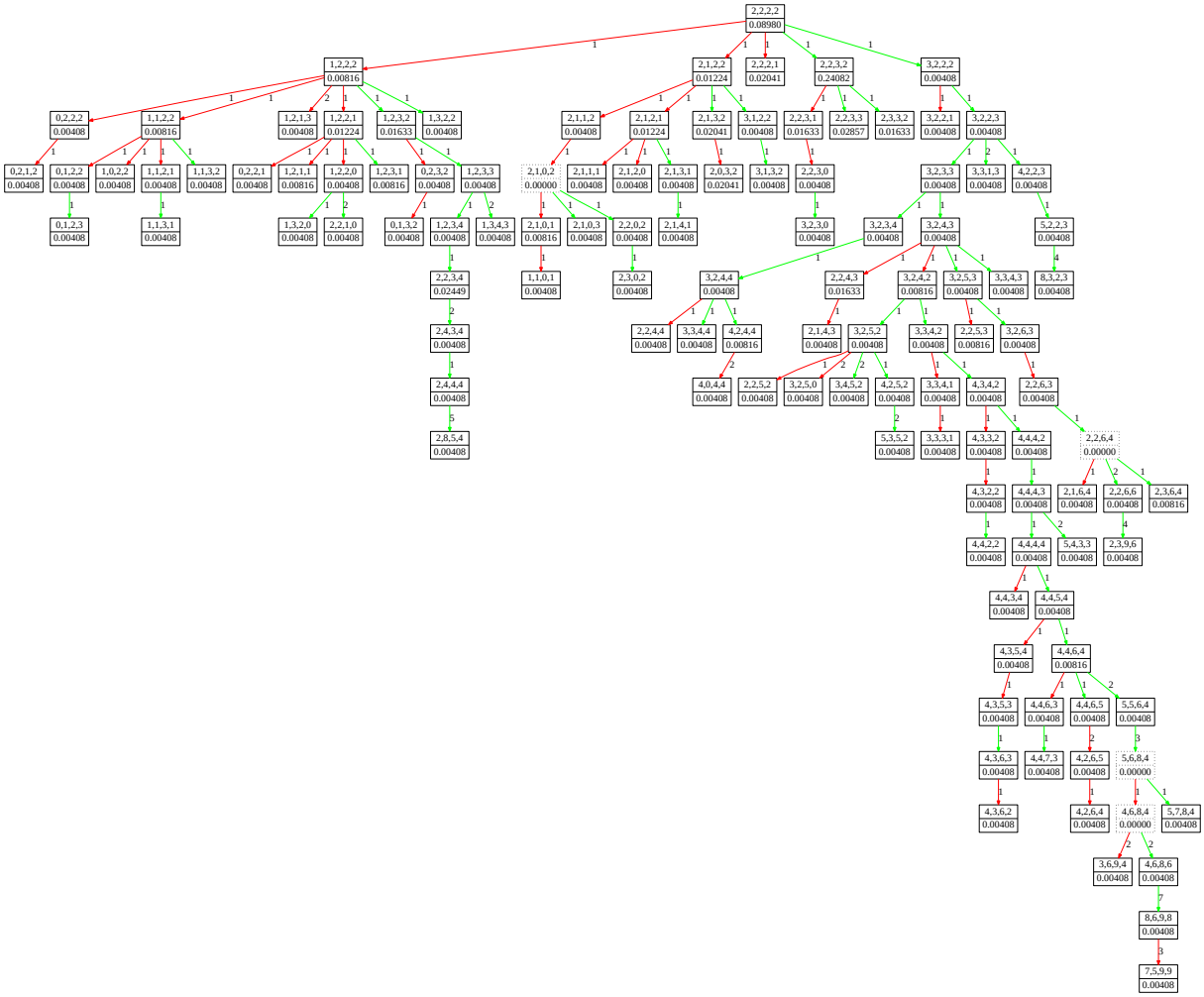

Figure SF52: Phylogenetic trees showing progression of primary stage cervical cancer in patient 31. The trees are built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the trees represents a copy number profile of the four gene probes *LAMP3, PROX1, PRKAA1* and *CCND1*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

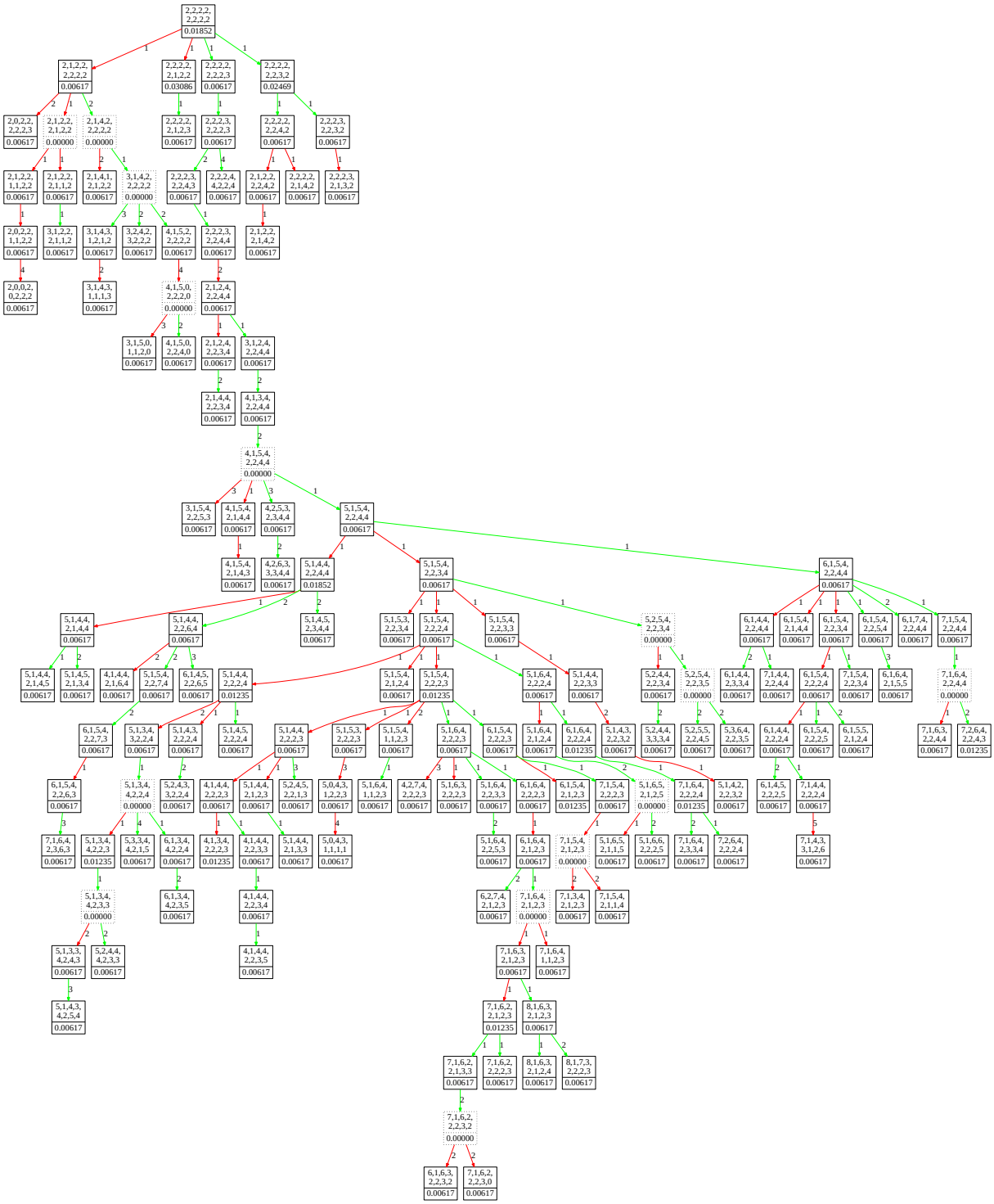

Figure SF53: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 1. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

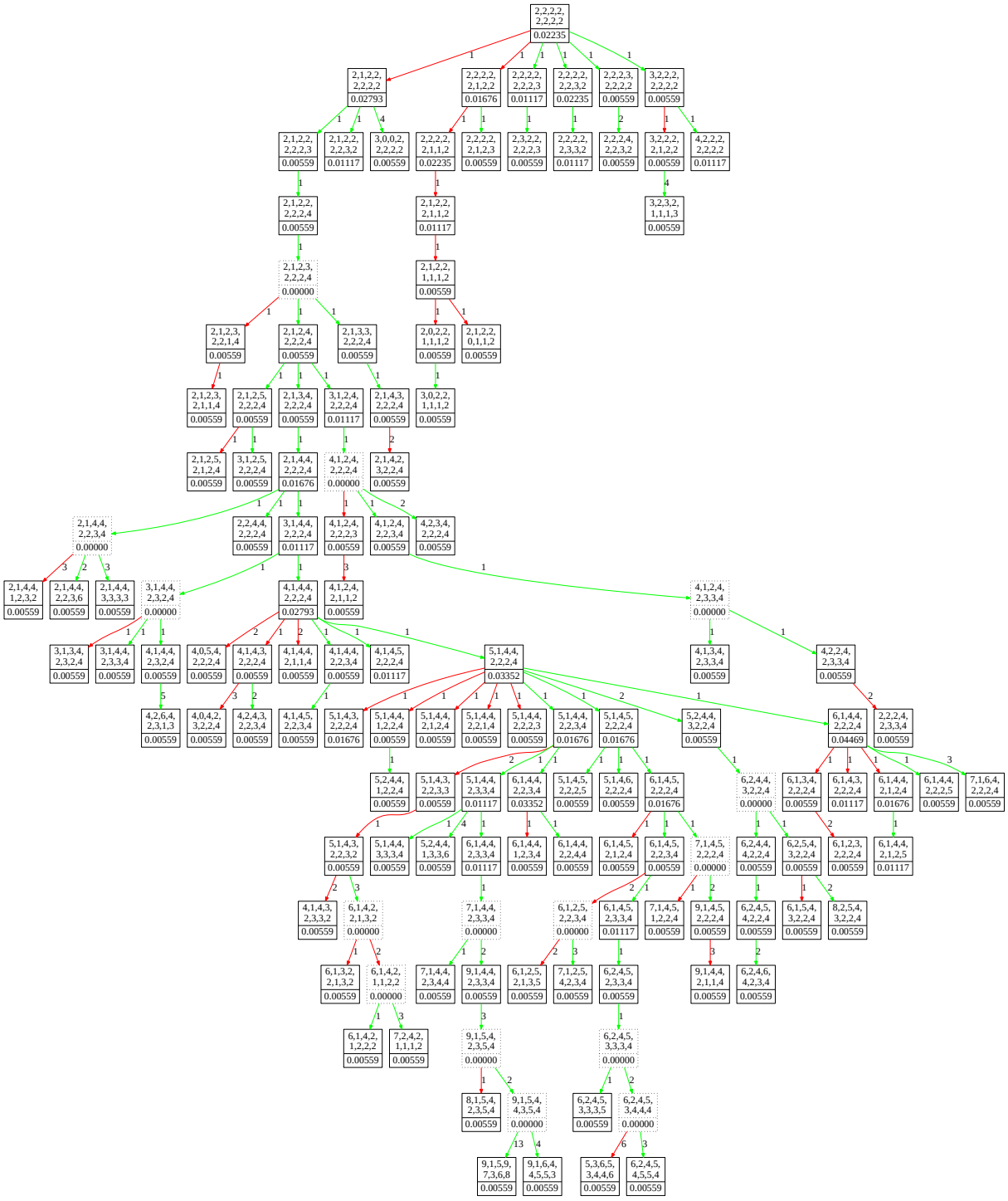

Figure SF54: Phylogenetic tree showing progression of IDC stage breast cancer in patient 1. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

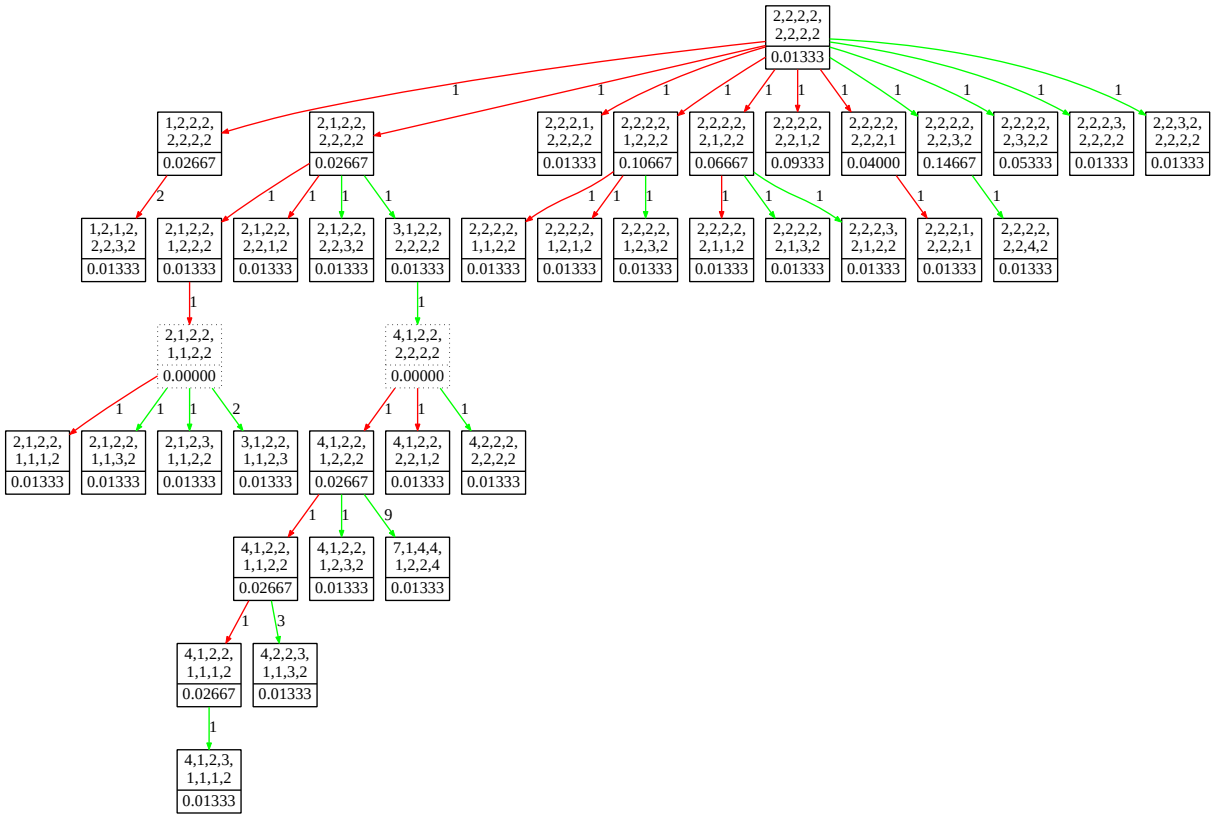

Figure SF55: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 2. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

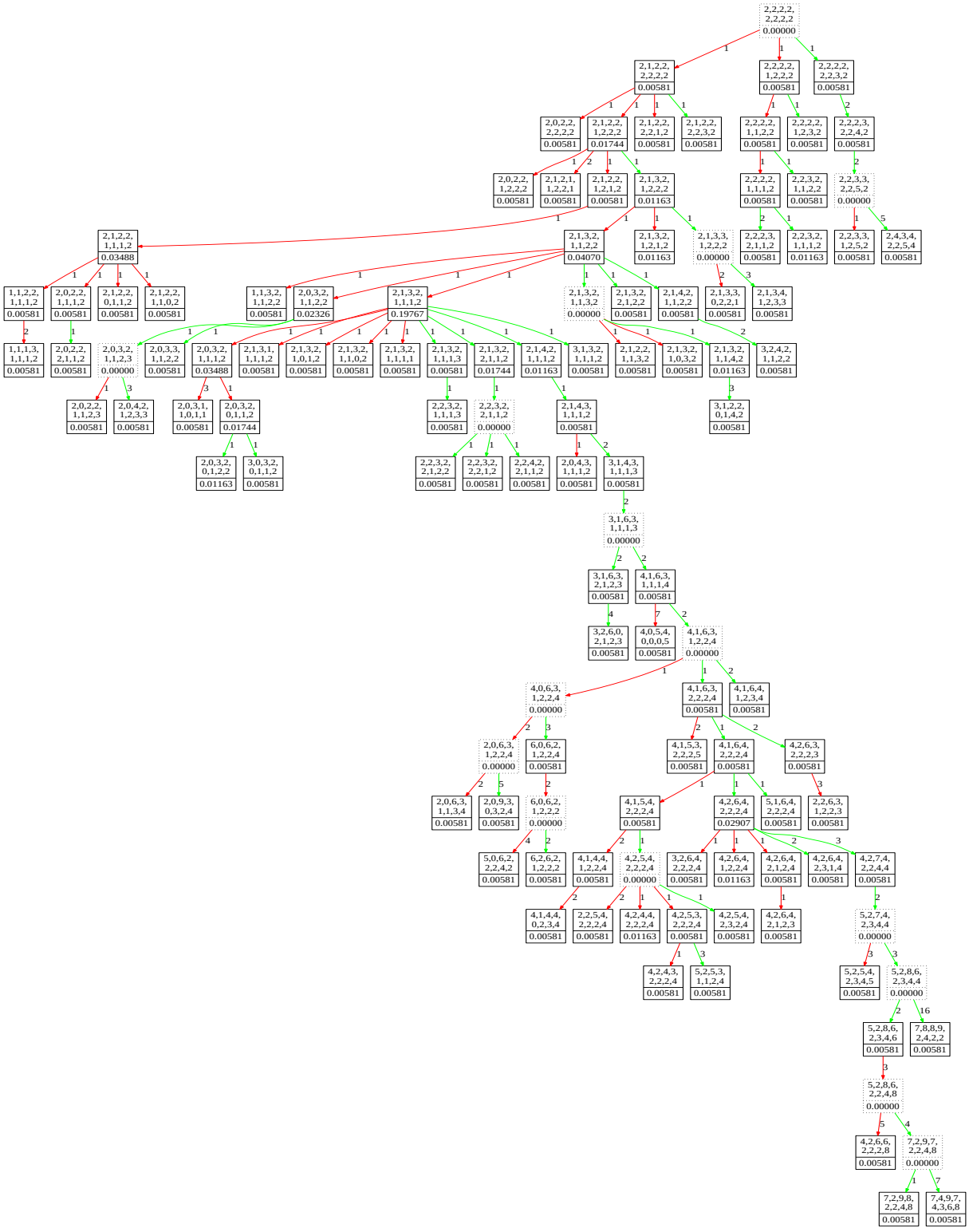

Figure SF56: Phylogenetic tree showing progression of IDC stage breast cancer in patient 2. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

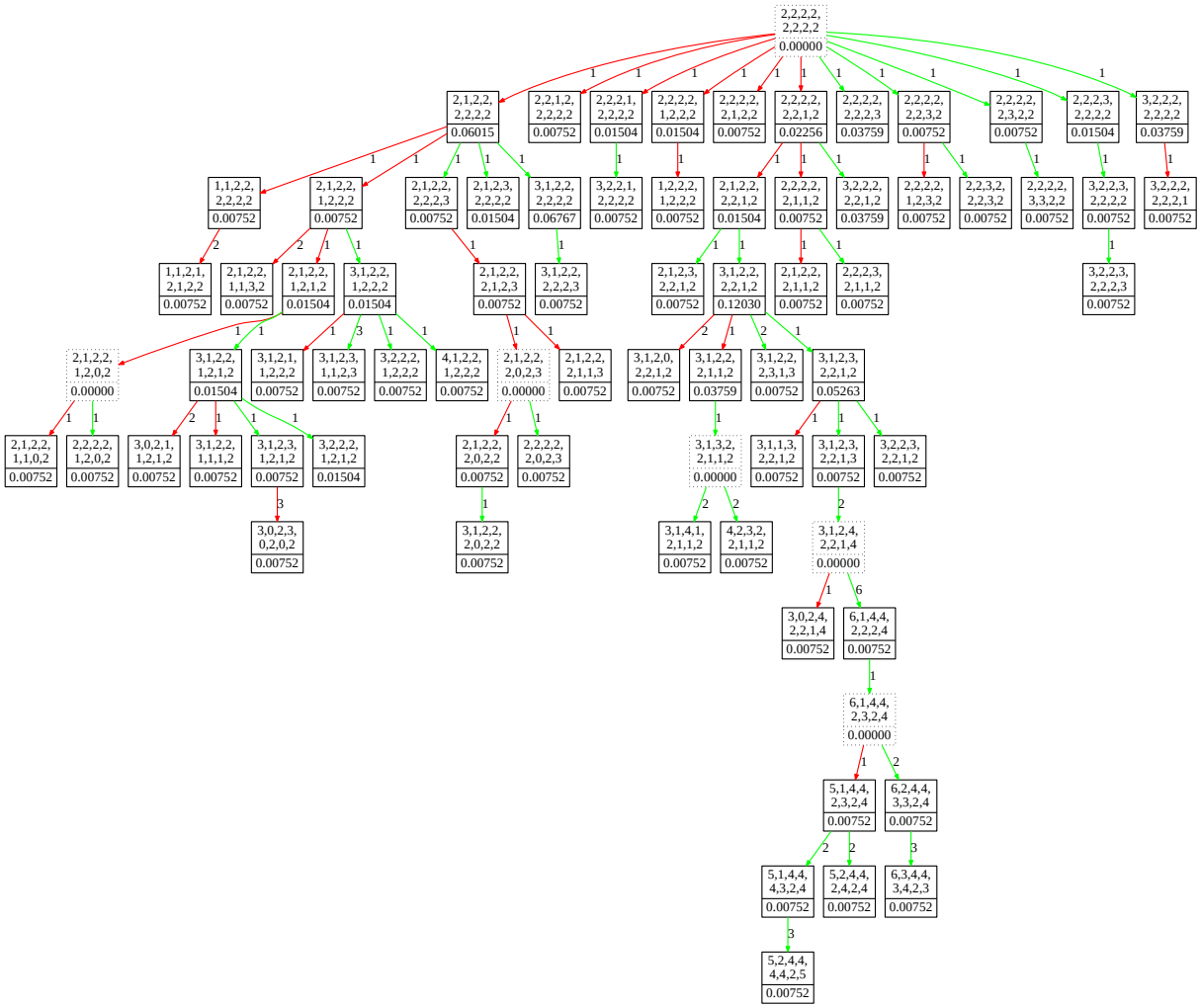

Figure SF57: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 3. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

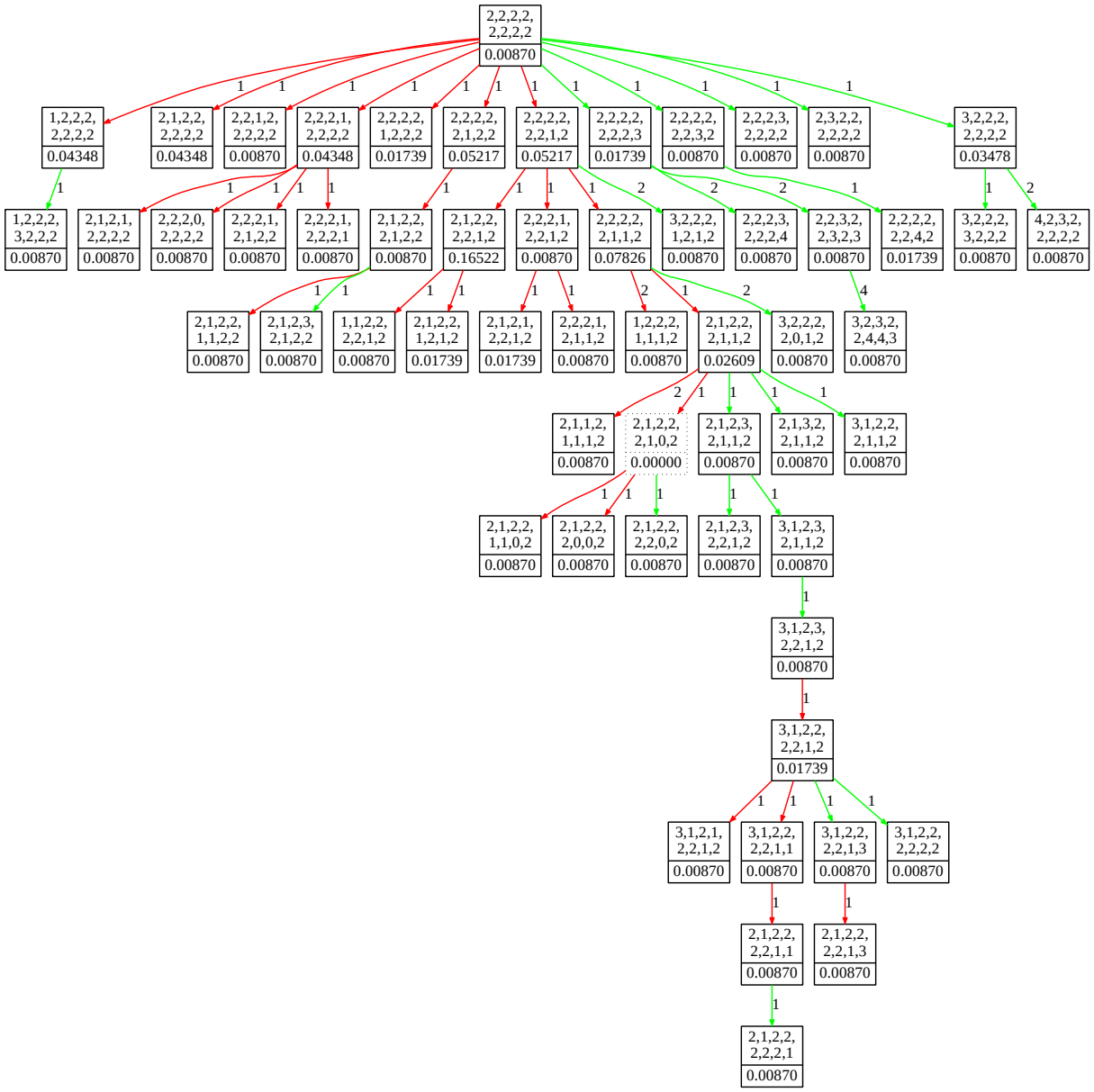

Figure SF58: Phylogenetic tree showing progression of IDC stage breast cancer in patient 3. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

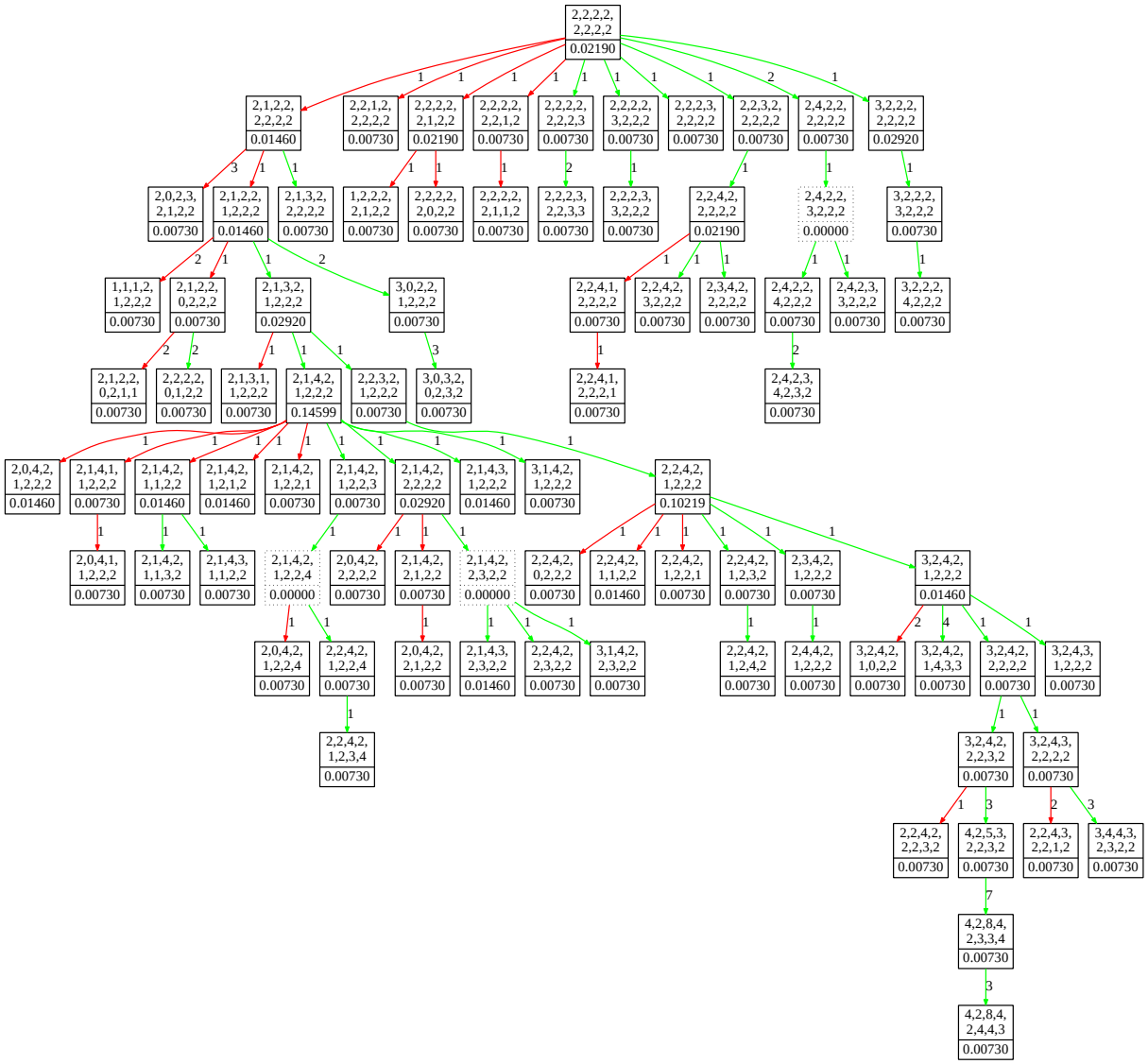

Figure SF59: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 4. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

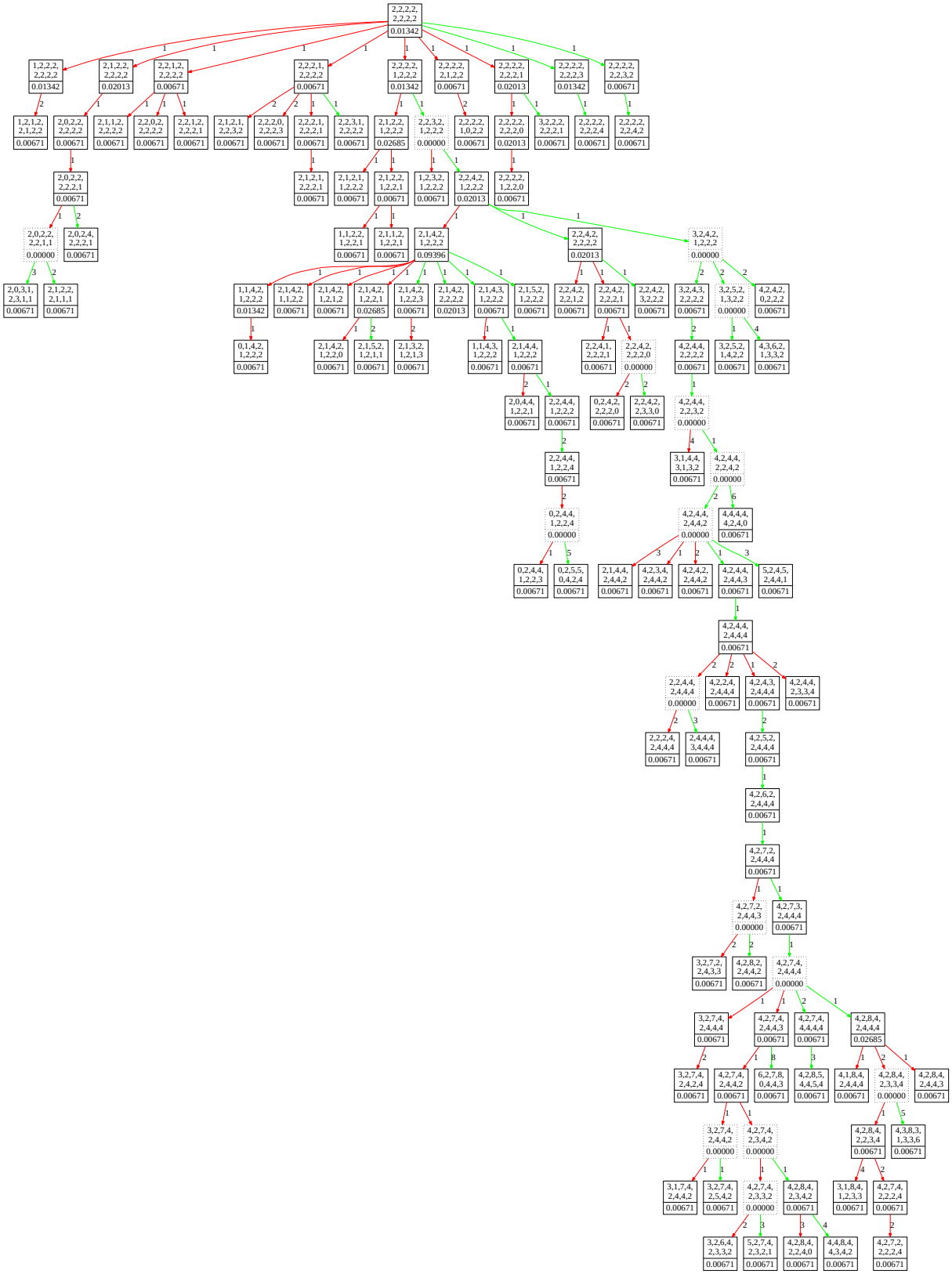

Figure SF60: Phylogenetic tree showing progression of IDC stage breast cancer in patient 4. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

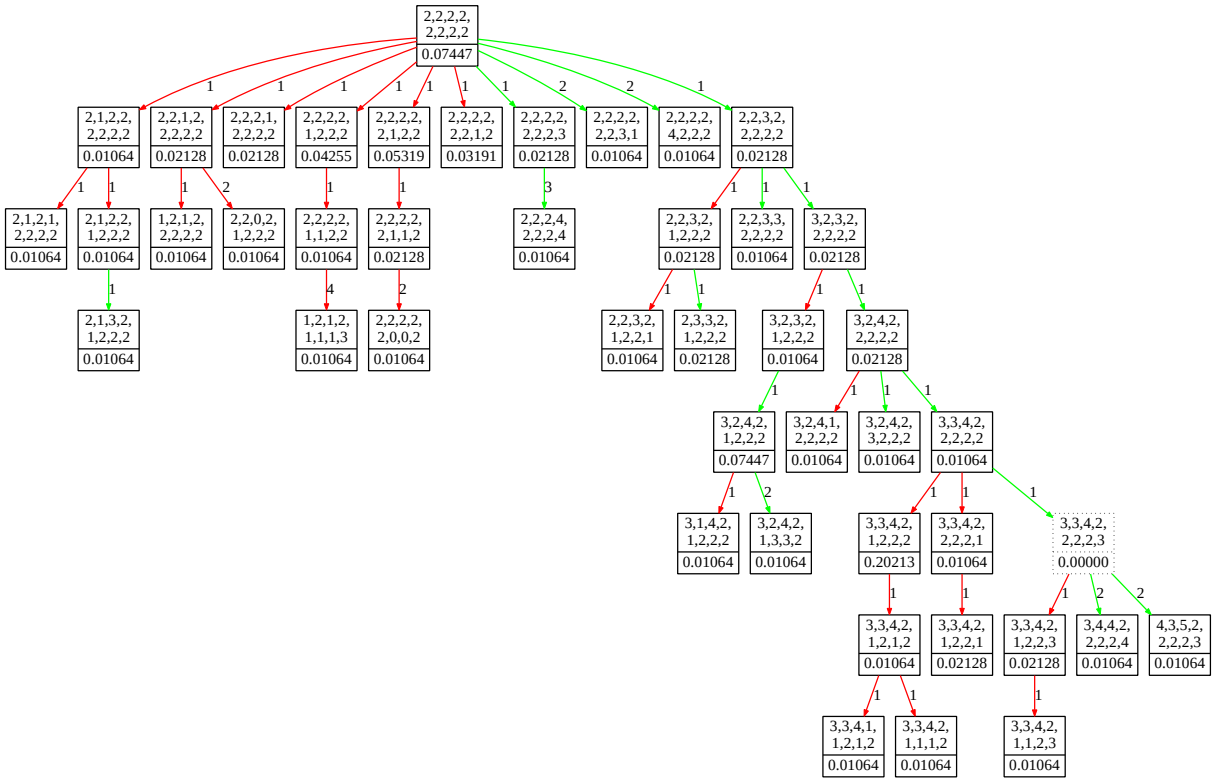

Figure SF61: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 5. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

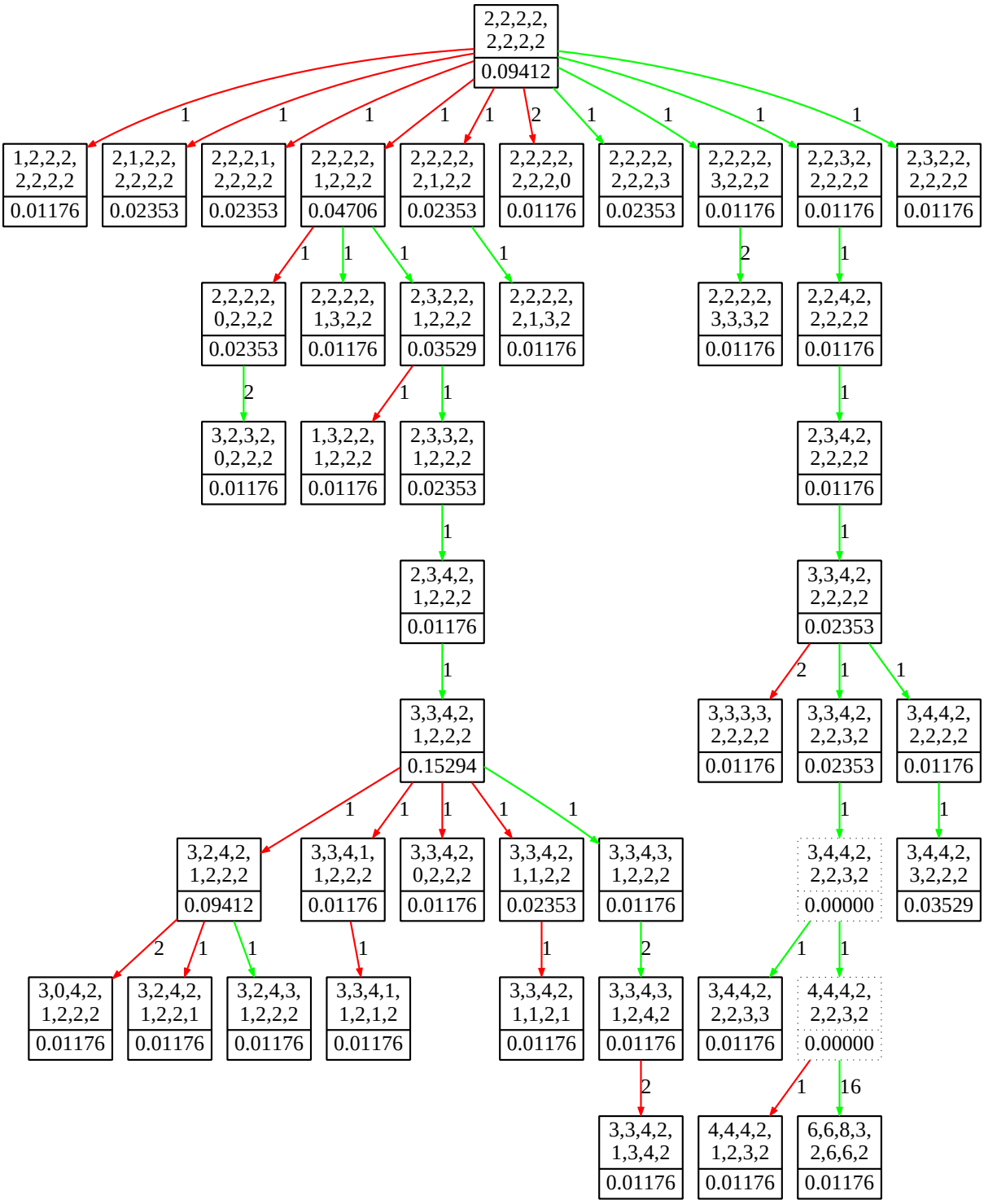

Figure SF62: Phylogenetic tree showing progression of IDC stage breast cancer in patient 5. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

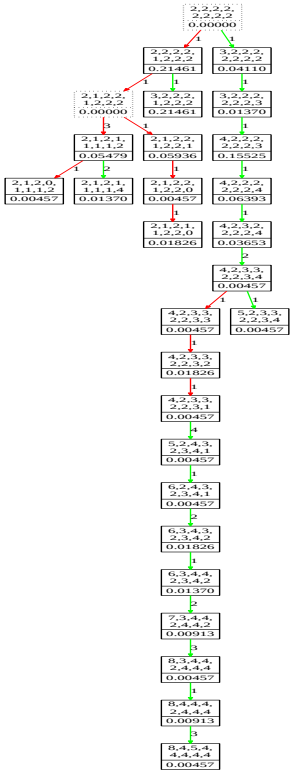

Figure SF63: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 6. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

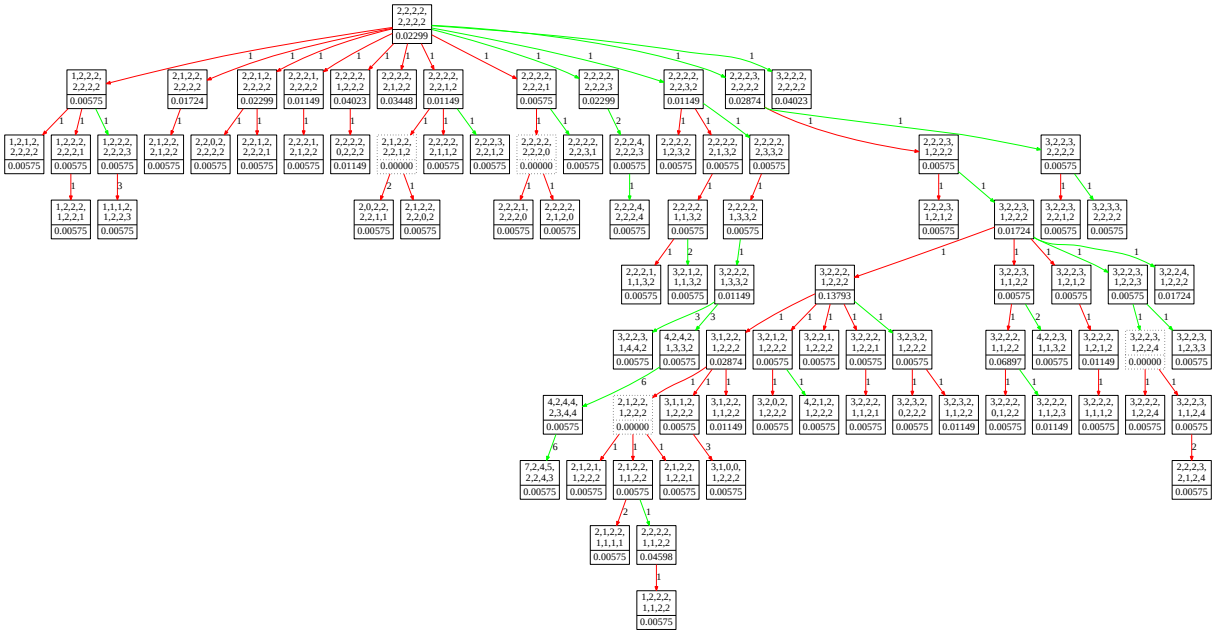

Figure SF64: Phylogenetic tree showing progression of IDC stage breast cancer in patient 6. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

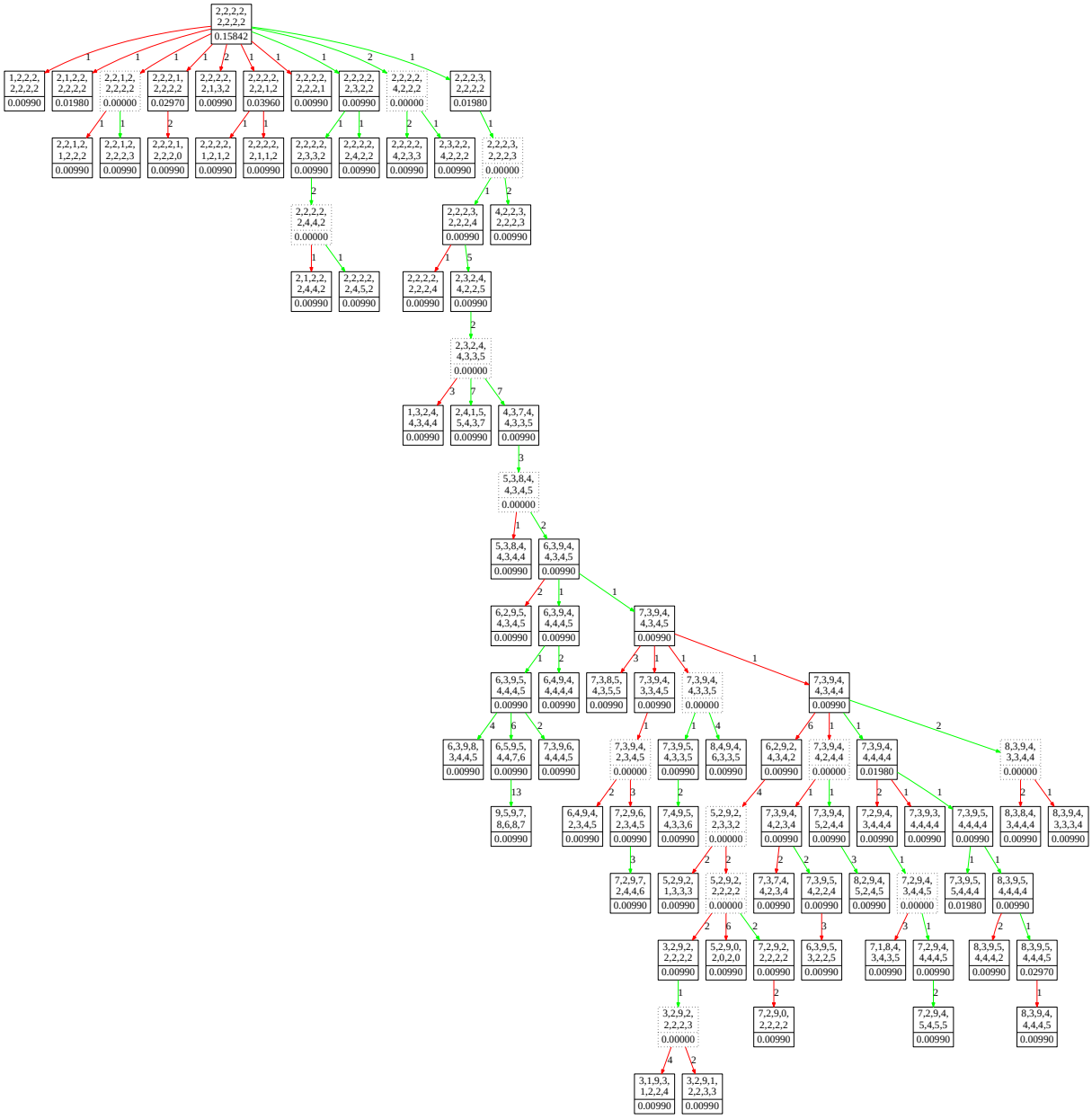

Figure SF65: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 7. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

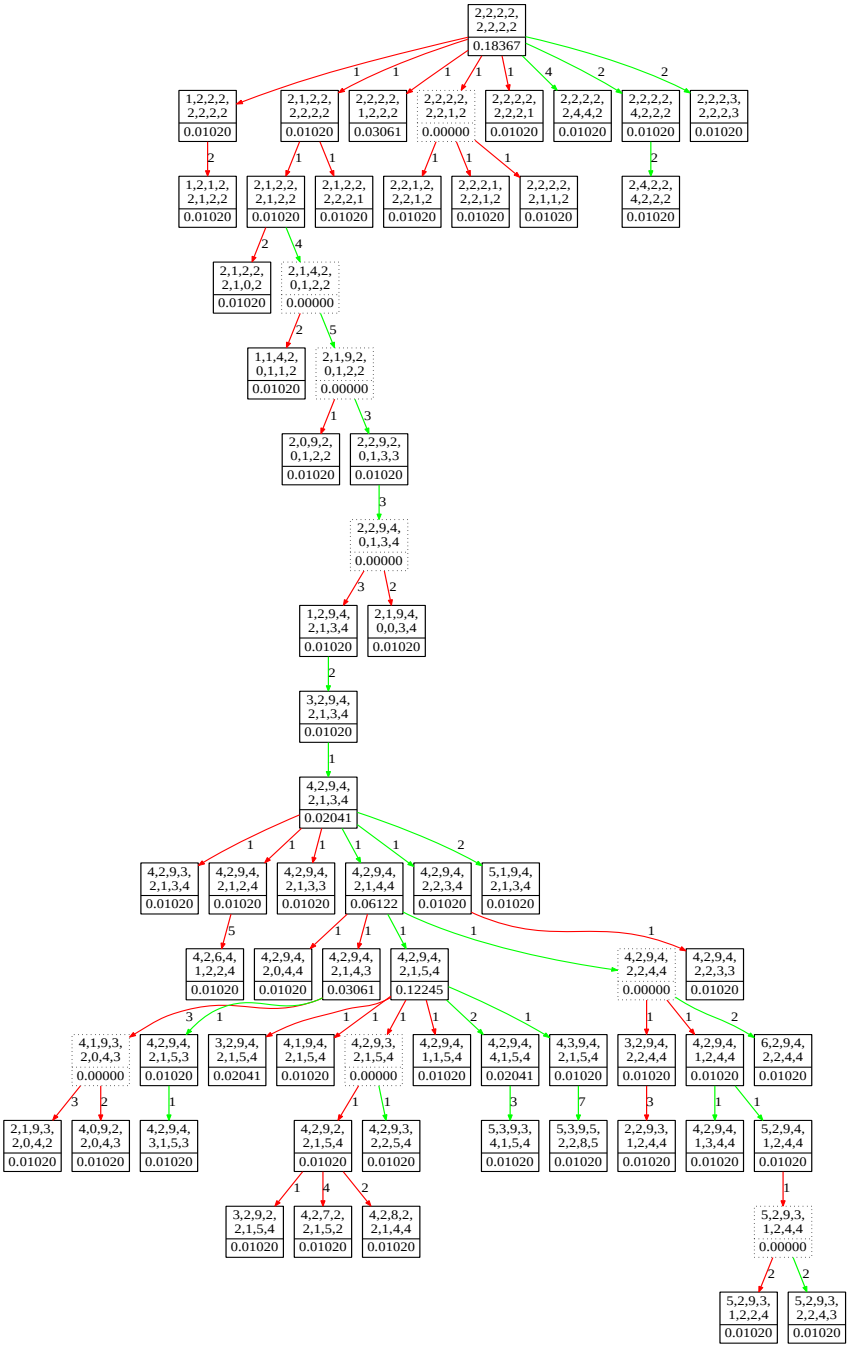

Figure SF66: Phylogenetic tree showing progression of IDC stage breast cancer in patient 7. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

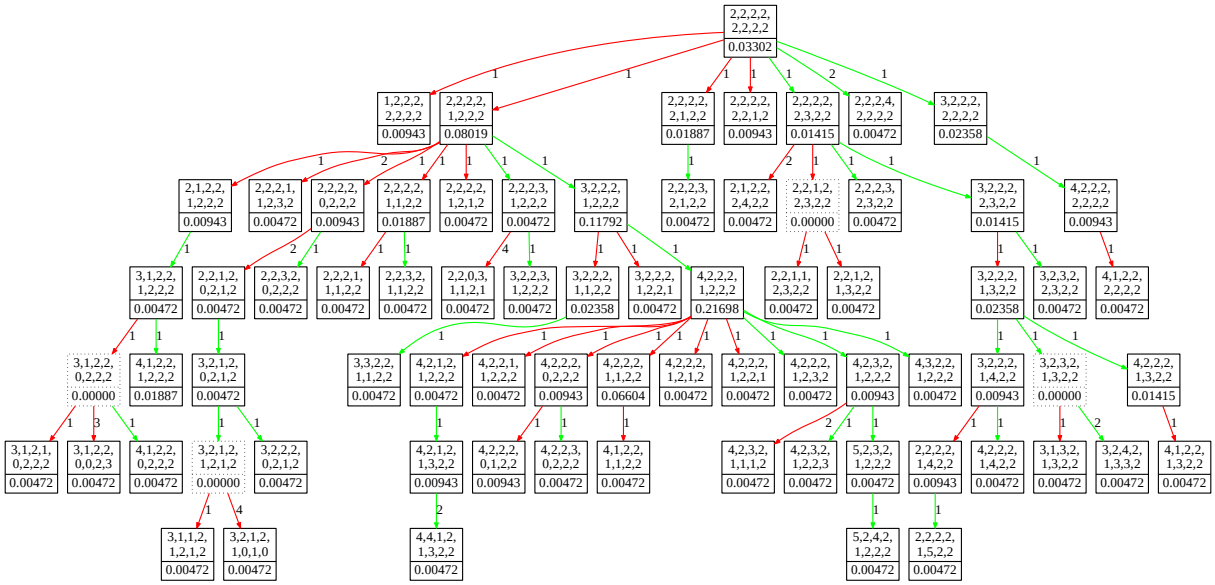

Figure SF67: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 8. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

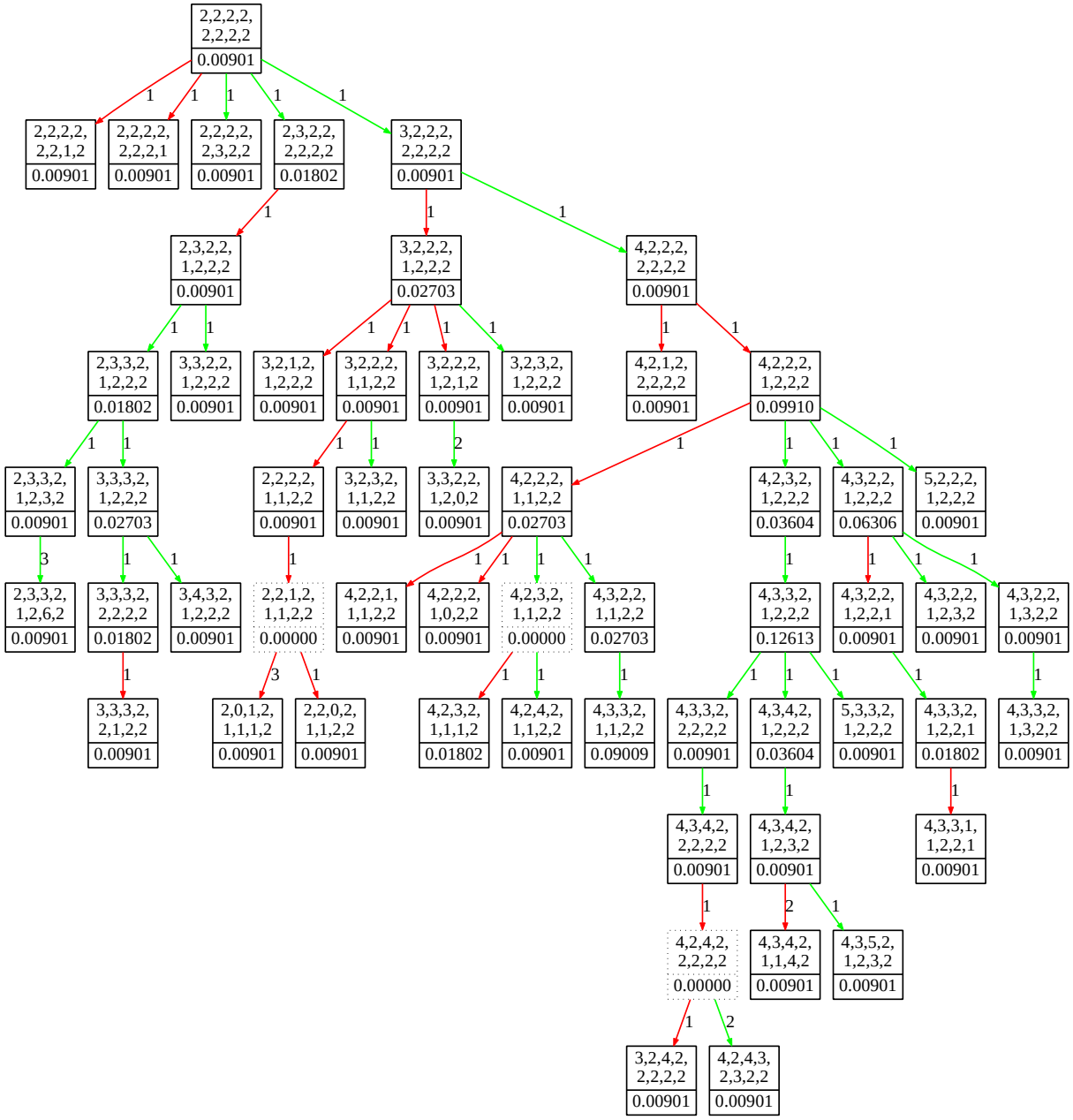

Figure SF68: Phylogenetic tree showing progression of IDC stage breast cancer in patient 8. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

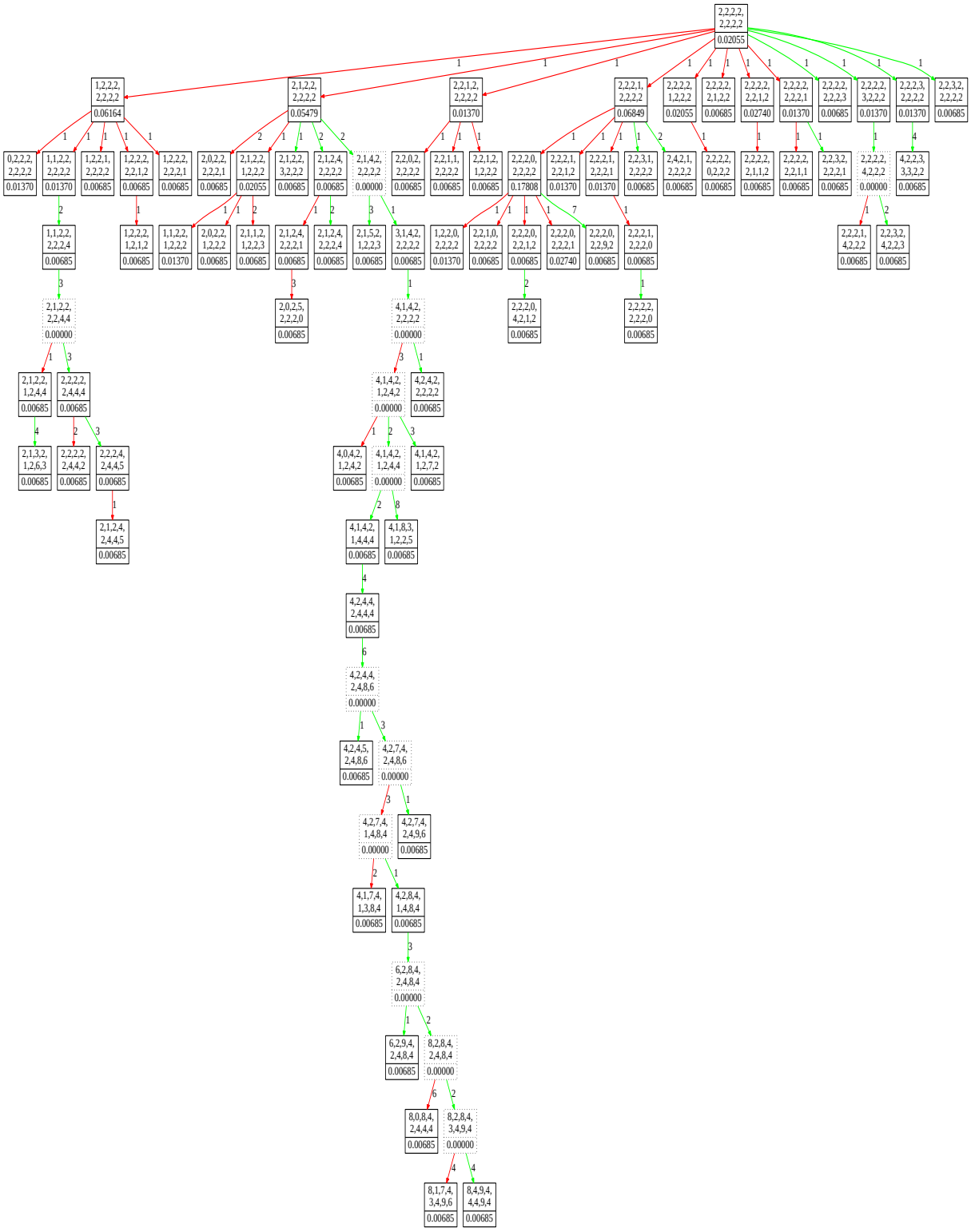

Figure SF69: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 9. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

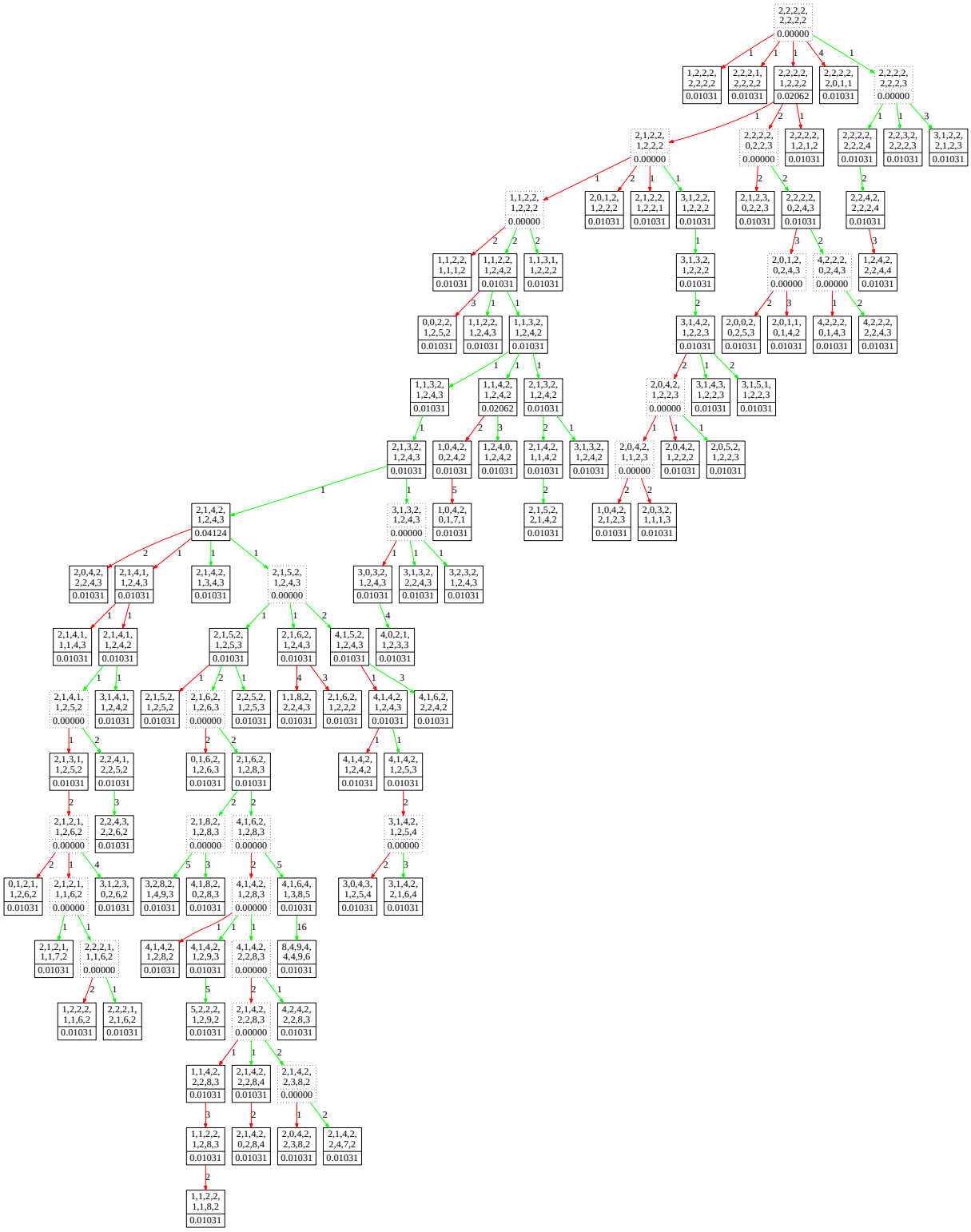

Figure SF70: Phylogenetic tree showing progression of IDC stage breast cancer in patient 9. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

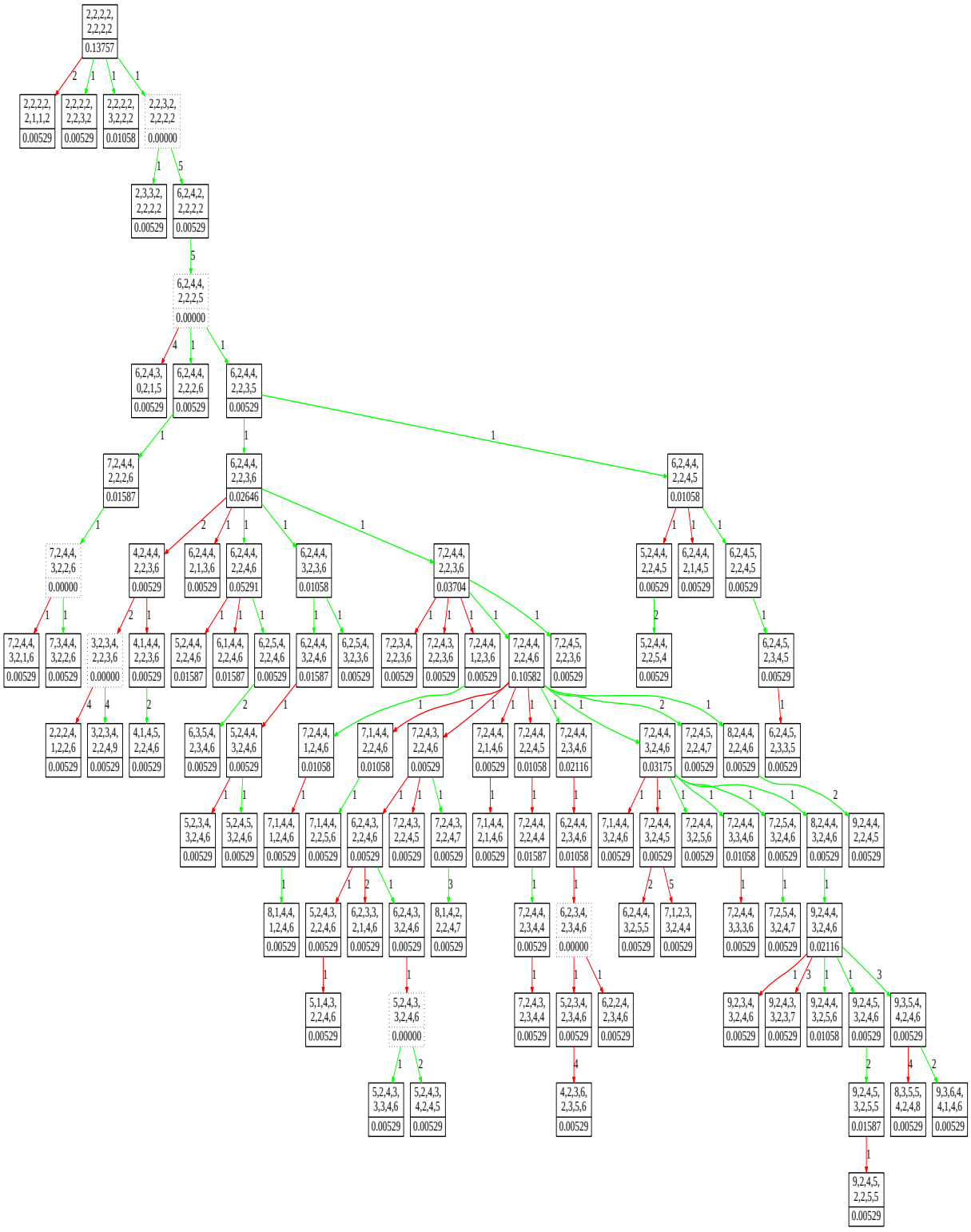

Figure SF71: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 10. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

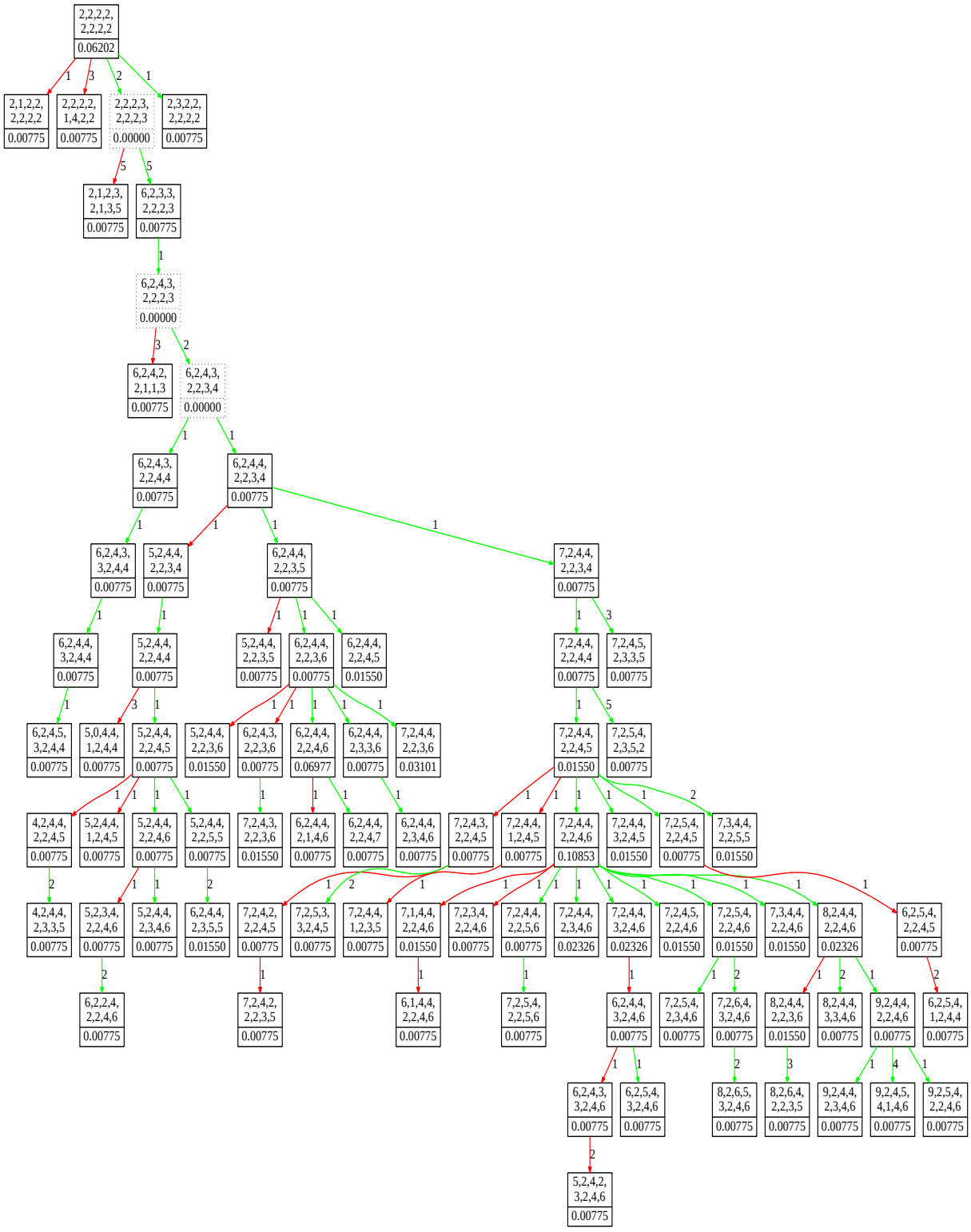

Figure SF72: Phylogenetic tree showing progression of IDC stage breast cancer in patient 10. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

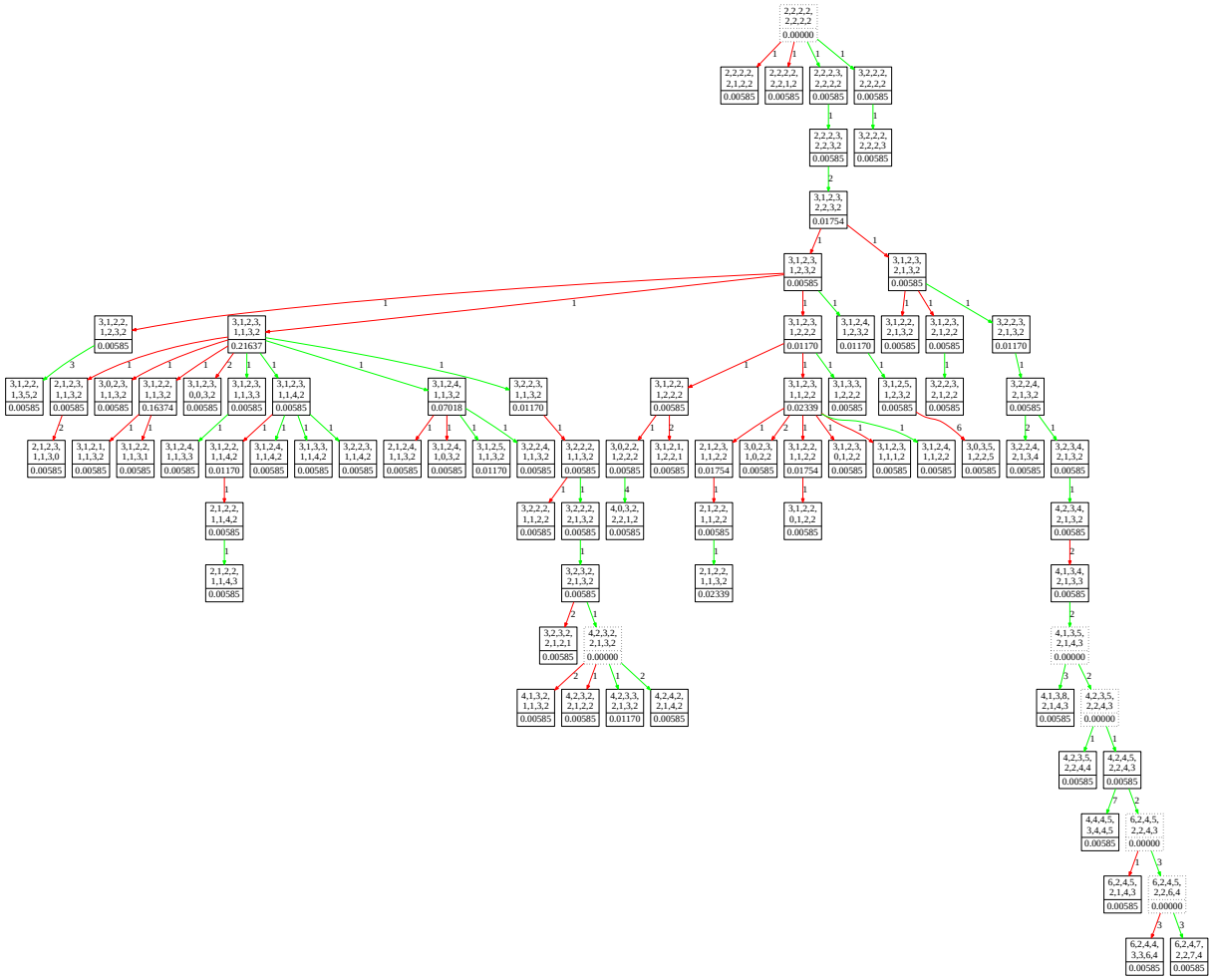

Figure SF73: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 11. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

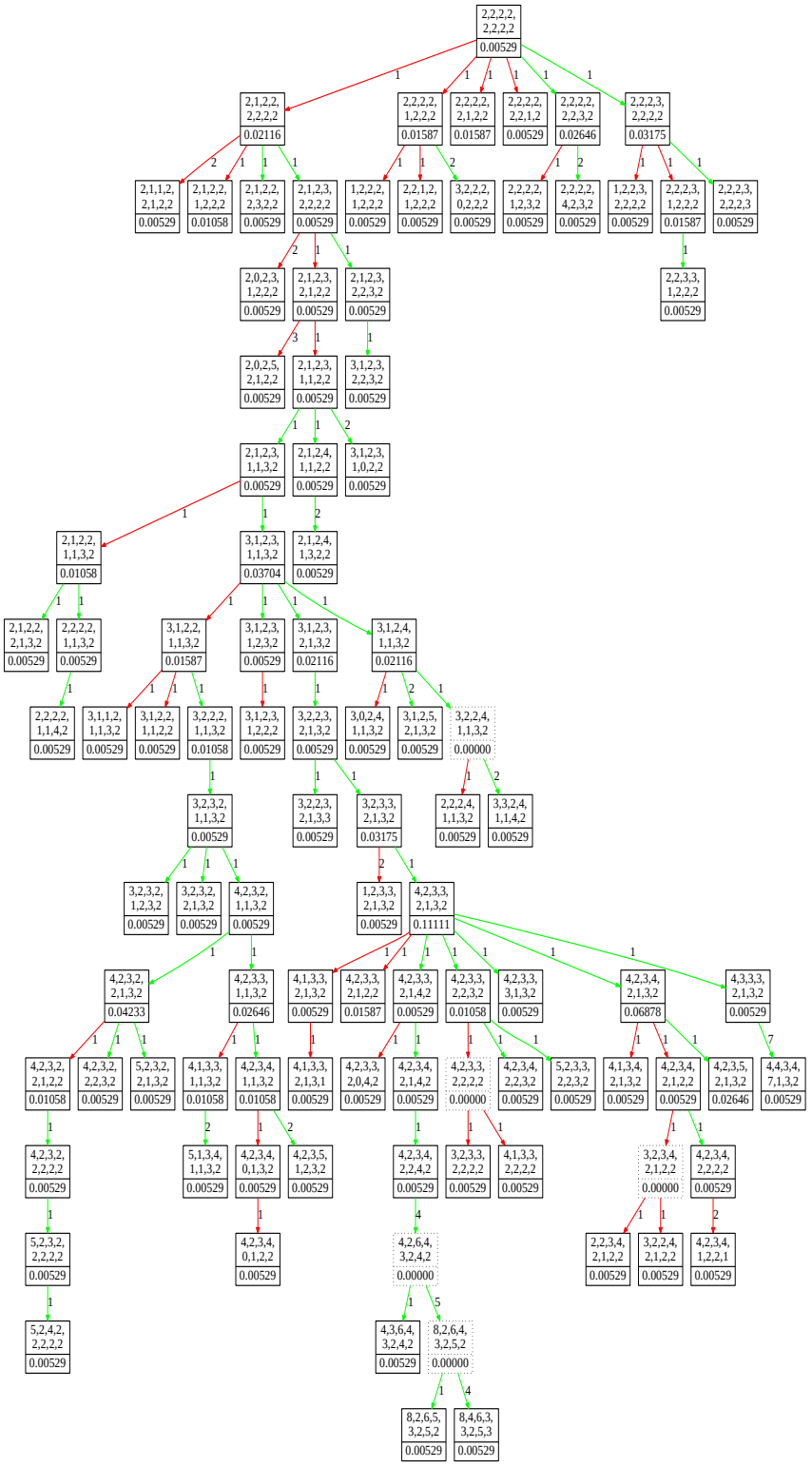

Figure SF74: Phylogenetic tree showing progression of IDC stage breast cancer in patient 11. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

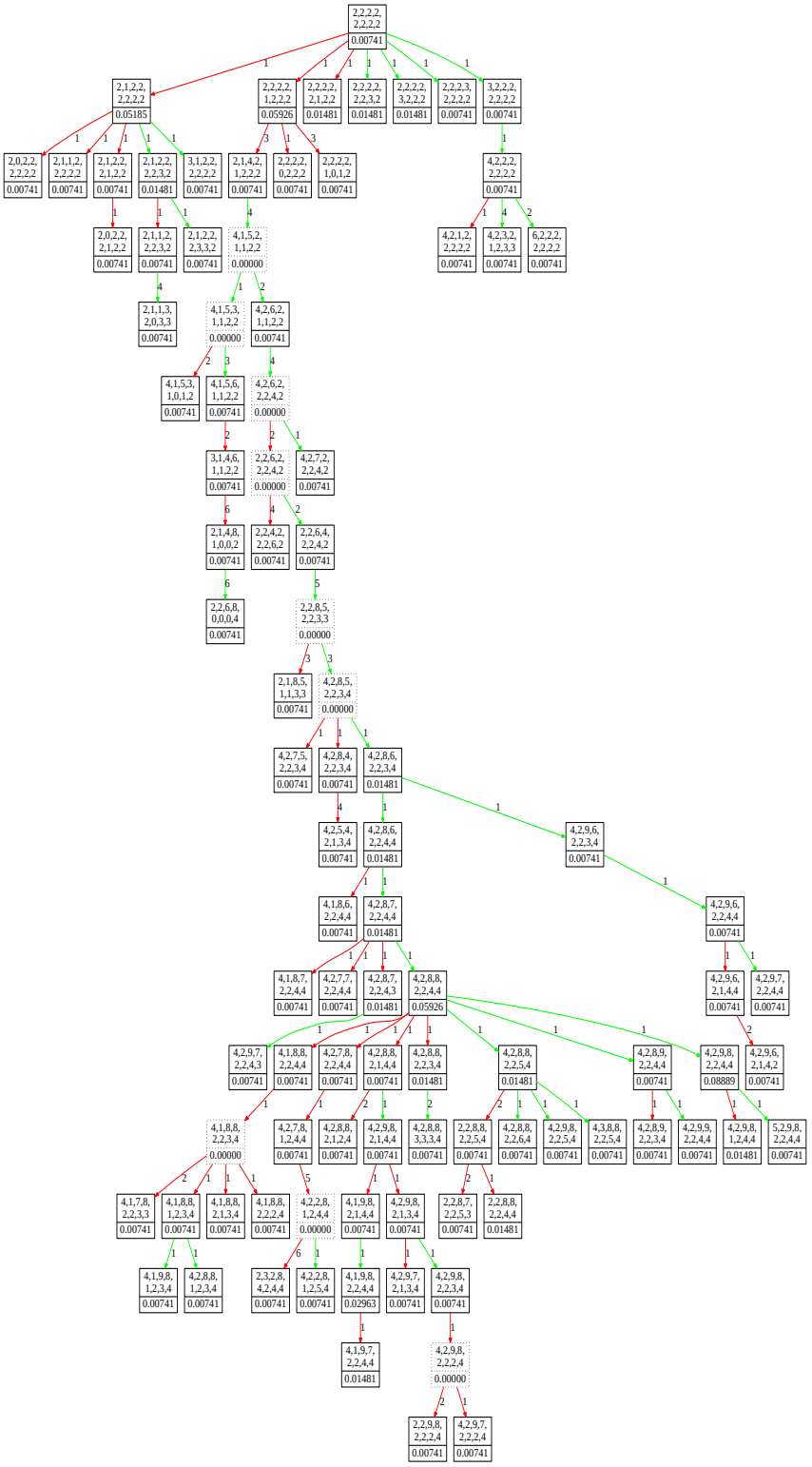

Figure SF75: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 12. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

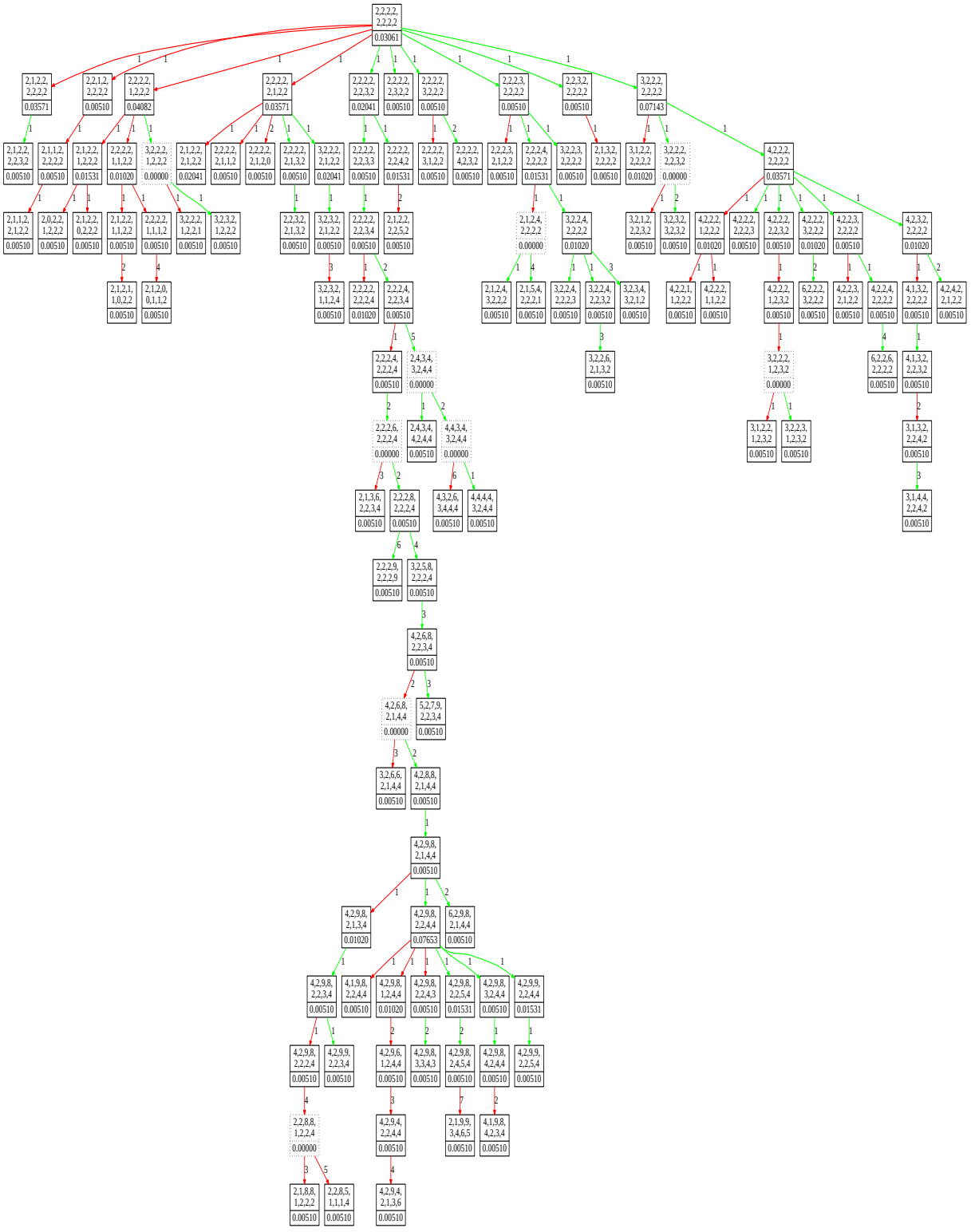

Figure SF76: Phylogenetic tree showing progression of IDC stage breast cancer in patient 12. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

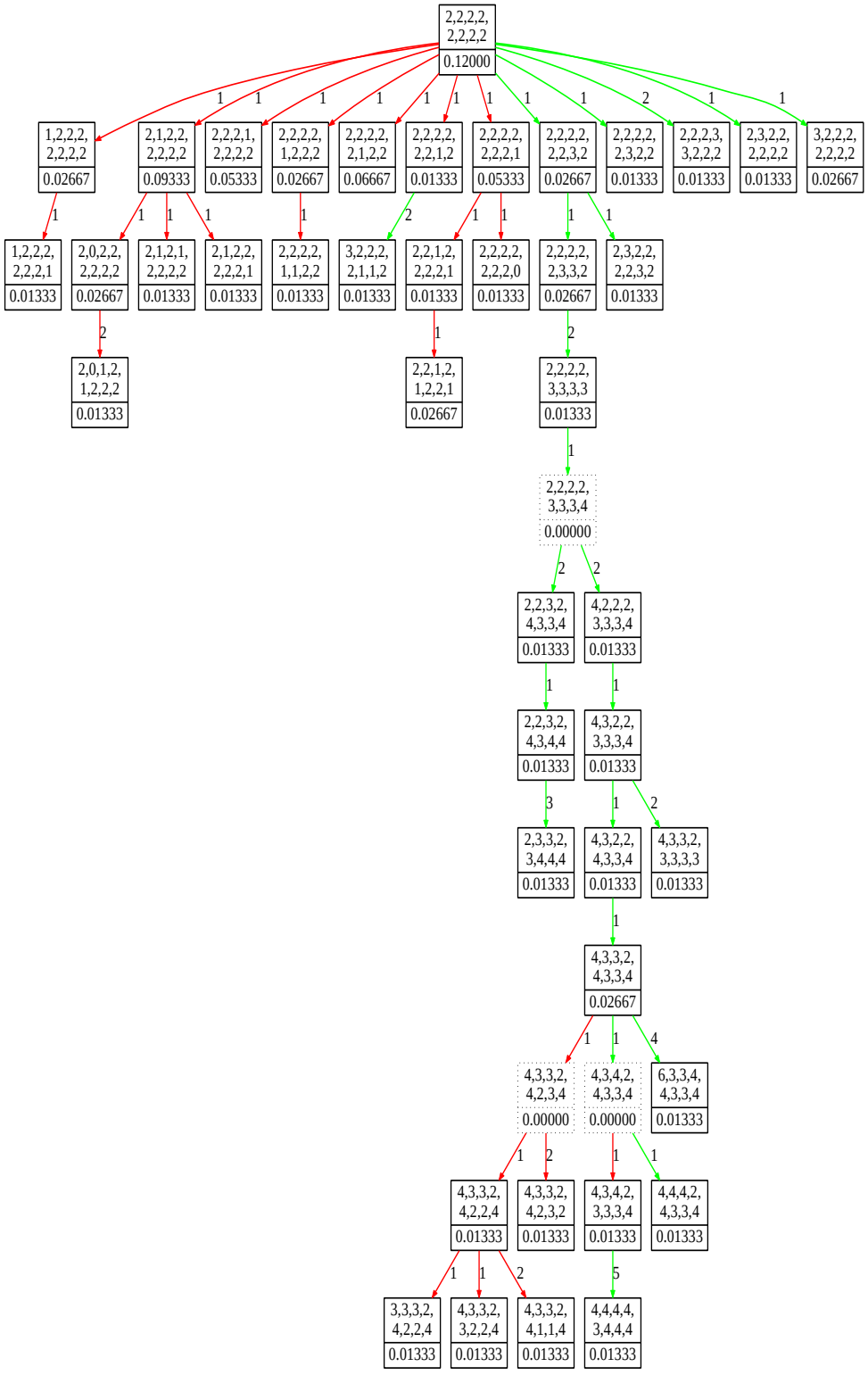

Figure SF77: Phylogenetic tree showing progression of DCIS stage breast cancer in patient 13. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.

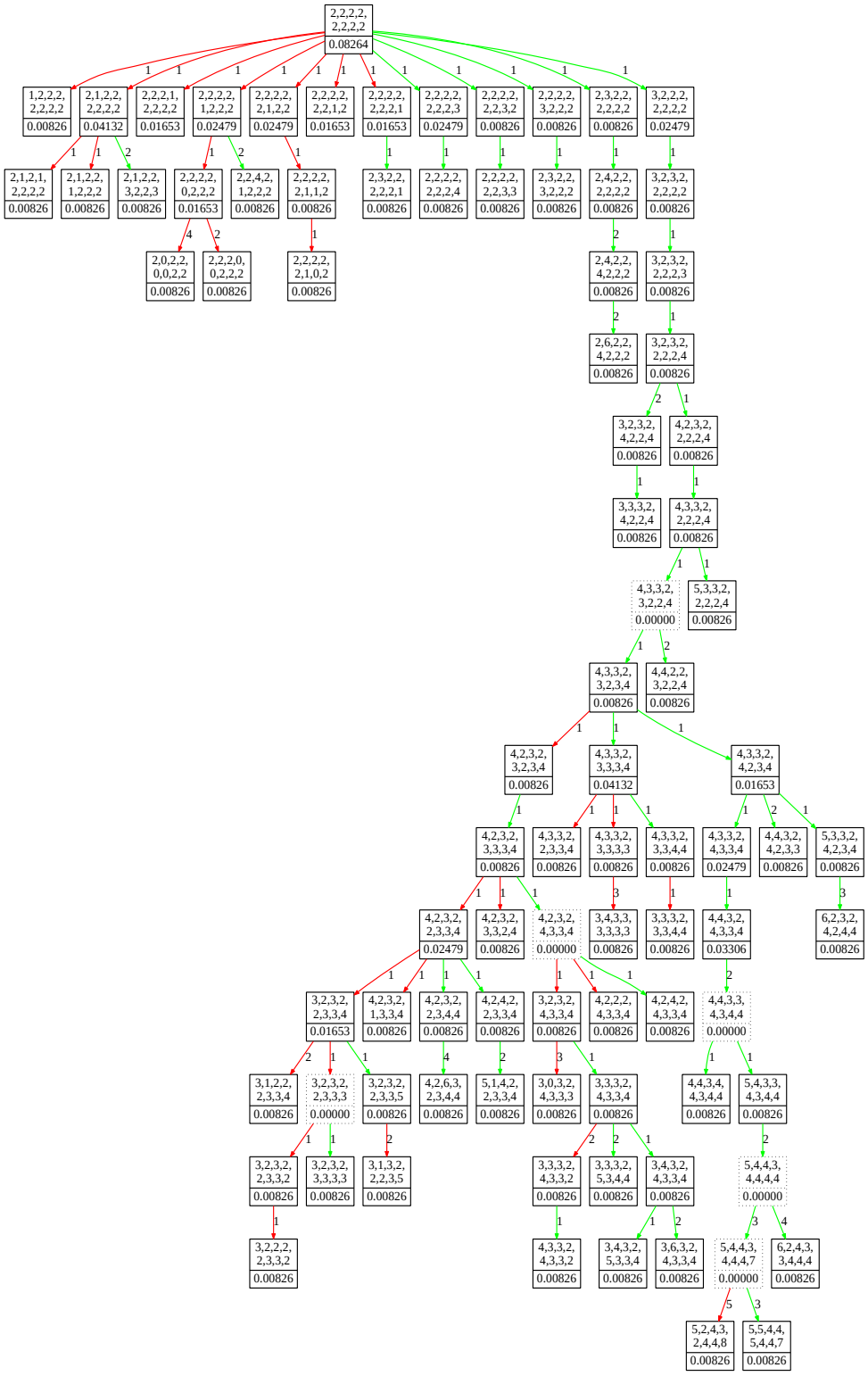

Figure SF78: Phylogenetic tree showing progression of IDC stage breast cancer in patient 13. The tree is built from single cell-copy number data using the ploidyless heuristic approach implemented in FISHtrees. Each node in the tree represents a copy number profile of the eight gene probes *COX-2, DBC2, MYC, CCND1, CDH1, p53, HER-2* and *ZNF217*. Nodes with solid borders represent cells present in the collected sample, while nodes with dotted borders represent inferred Steiner nodes. Green and red edges model gene gain and gene loss respectively. The weight value on each edge connecting two nodes x and y is the rectilinear distance between the states of x and y. The weight on each node describes the fraction of cells in the sample with the particular copy number profile modeled by that node; Steiner nodes are assigned weight 0.