Genetic bottlenecks and the hazardous game of population reduction in cell line based research

David Gisselsson,[a*] David Lindgren,[b] Linda Holmquist Mengelbier,[a] Ingrid Øra,[c] and Herman Yeger[d]

[a]Department of Clinical Genetics, Lund University, University and Regional Laboratories, Lund, Sweden

[b]Department of Laboratory Medicine, Center for Molecular Pathology, Lund University, Malmö, Sweden

[c]Department of Pediatric Oncology, Skåne University Hospital, Lund, Sweden

[d]Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada.[a]

Keywords: cell line; cancer; genetic bottle neck; population bottle neck; SNP-array; chromosomal instability

*Corresponding author. David Gisselsson, Department of Clinical Genetics, Lund University, University Hospital, SE 221 85 Lund, Sweden. Phone +46 46 173418. Fax +46 46 13 10 61. E-mail address: david.gisselsson_nord@med.lu.se

Established tumour cell lines are ubiquitous tools in research, but their representativity is often debated. One possible caveat is that many cell lines are derived from cells with genomic instability, potentially leading to genotype changes *in vitro*. We applied SNP-array analysis to an established tumour cell line (WiT49). Even though WiT49 exhibited chromosome segregation errors in 30% of cell divisions, only a single chromosome segment exhibited a shift in copy number after 20 population doublings in culture. In contrast, sub-populations derived from single cells expanded for an equal number of population doublings showed on average 5.8 and 8.9 altered segments compared to the original culture and to each other, respectively. Most copy number variants differentiating these single cell clones corresponded to pre-existing variations in the original culture. Furthermore, no sub-clonal variation was detected in any of the populations derived from single cells. This indicates that genetic bottlenecks resulting from population reduction poses a higher threat to genetic representativity than prolonged culture *per se*, even in cell lines with a high rate of genomic instability. Genetic bottlenecks should therefore be considered a potential caveat in all studies involving sub-cloning, transfection and other conditions leading to a temporary reduction in cell number.

Introduction

Established tumour cell lines are common tools in cancer research and drug discovery, but their representativity of *in vivo* tumour disease has remained an issue of much debate [1,2]. It is well known that the extent to which cancer cell lines recapitulate the features of their original tumours is highly variable with respect to morphology, gene expression, and genetic alterations [3]. Furthermore, many cell lines are derived from tumours with an inherent genomic instability [4-8] that may potentially lead to continuous evolution of novel genetic features during prolonged *in vitro* growth. Cytogenetic studies have demonstrated the presence of two stages of karyotype evolution during the establishment of cancer cell lines, an early phase characterized by cytogenetic heterogeneity and selection of clones fit for *in vitro* propagation, and a later phase which is relatively more stable with respect to chromosomal alterations [9]. A number of studies have been performed to compare genetic alterations in cell lines with those present in the corresponding primary tumours. For example, in breast cancer it has been shown that aberrations found in primary tumours are typically retained after cell line establishment [10] while a recent high-resolution genomic characterization of colorectal cancer cell lines has shown that the general pattern of genomic alterations is concordant with the panorama of cytogenetic alterations in primary tumour material. However, the cytogenetic features of cell lines and primary tumours are still not completely similar, indicating that clonal evolution is to some extent maintained *in vitro* [11]. This is accentuated by the fact that sub-lines derived from the same tumour may exhibit distinct genomic profiles. Detailed analyses of different sub-lines from the commonly used MCF-7 breast cancer cell line have shown that there is wide genetic variation among the different lines with respect to genomic imbalances and expression profiles [12]. This considered, surprisingly little is known about the principles guiding *in vitro* genetic evolution of established cancer cell lines after the first phase of clonal selection at cell line establishment.

This issue is nevertheless important because many established cell lines are intensely used in research projects in which they may be subjected to sub-culturing, transfection, infection, starvation, and other changes in growth conditions after which sub-populations having undergone different biological manipulations are typically compared to each other without taking into account potential genetic differences between these sub-populations.

We envisioned two principally different scenarios by which established cancer cell lines may change in genotype during *in vitro* growth (Fig. 1). Both these models assume that the original tumour cell population has some degree of genomic instability and is therefore genetically heterogeneous, consisting of more than one sub-clone that can be defined according to its genomic profile. This is consistent with the situation found in the majority of commonly used human cell lines [4-8]. Under this assumption, the capacity for formation of clones with novel chromosome aberrations *in vitro* will depend largely on the fitness of cells having acquired novel genomic changes. Based on this, two diametrically opposite scenarios can be envisioned. First, the proliferative survival of cells with novel genetic aberrations can be sufficiently high to allow them to expand and compete with previously formed cell populations, resulting in continuous change of a cell line's genotype during *in vitro* growth. In essence, this implies that the fitness of cells with novel aberrations compared to cells not having acquired novel genetic changes is higher or equivalent. We tentatively attributed the term *clonal liberalism* to this model, as it reflects a situation where novel clones compete with earlier clones in a disorganized fashion. In the other scenario, the fitness is close to 0 for cells with novel genomic aberrations, implying that clones with novel changes are rarely formed. During prolonged culture *in vitro*, variability in genotype of the bulk cell line population will therefore be limited to changes in the prevalence of already pre-existing clones. We attributed the term *clonal conservatism* to this model. Cell lines functioning according to the *clonal*

conservatism model may, however, still be subject to ostensibly radical changes in genotype by passing through a genetic bottleneck (also referred to as a population bottleneck), *i.e.* situations when population numbers are temporarily reduced to a level insufficient to maintain the diversity in the population. Most genomic screening techniques available today are unable to detect chromosome aberrations present in cell populations of low prevalence and the genomic profile of a cell line's bulk population will be dominated by the clones having the highest prevalence. However, in any situation when the proliferating population is reduced to a very small number of cells, any genetically distinct sub-population that normally would have a prevalence level too low for detection by genomic screening of the original bulk cell line can potentially re-grow into a new population with a different genomic profile than the original bulk. Genetic bottlenecks could be brought about by any situation when a minority cell population is allowed to expand to replace other cell populations of the cell line, for example by sub-cloning a small number of cells or in situations of extensive cell death or proliferation arrest. *Clonal liberalism* and *clonal conservatism* reflect two extreme situations, and it cannot be excluded that most tumour cell lines adhere to neither of these extremes. However, by using a cell line with high rate of chromosome segregation errors it should still be possible to evaluate which of the two models that is closest to the actual *in vitro* situation. The present study is an attempt to explore this issue using the WiT49 cell line.

Materials and Methods

Cell culture and cytogenetic

WiT49 was derived from a first-generation nude mouse subcutaneous xenograft of a lung metastasis of an aggressive primary nephroblastoma and has been carried in continuous culture for more than 4 years and through numerous passaging [13]. It has a hypotriploid karyotype with multiple complex structural and numerical chromosome aberrations and

exhibits extensive intercellular cytogenetic heterogeneity [14,15], similar to the vast majority of established cell lines used for research [9]. WiT49 was selected for the present study because it has a high frequency of chromosome segregation errors that could potentially cause genomic alterations in culture (see calculations below). Cells were cultured in RPMI medium supplemented with 10% fetal bovine serum, 100 IU/ml penicillin V and 100 µg/ml streptomycin. The average population doubling time was determined by holographic microscopic video capture for 3-8 days of ten samples consisting of 10-50 cells growing at 37 °C in ibidi µ-Slide I channel slides (ibidi, Martinsried, Munich, Germany), allowing continuous gas exchange through a plastic membrane. Sub-culturing was performed by trypsin treatment according to standard protocols. Single cell-derived clones were created by seeding of single cells in a Petri dish, after which 20 attached single cells were identified and monitored daily. After 72 h, the five largest colonies formed by these single were transferred by micro-pipette to separate wells in microtiter plates and expanded for 7 days, after which they were transferred to plastic culture flasks. All cell populations derived from single cells exhibited continuous log-phase growth during the period of expansion and were successfully harvested for DNA extraction after 30 days. Chromosome banding analyses and fluorescence in situ hybridization (FISH) on metaphase cells were according to standard methods [16], the latter by using commercially available whole chromosome painting probes (Abbott Molecular, Abbott Park, IL). Assessment of chromosome missegregation was performed as described in Gisselsson et al. [17] using a pan-alpha satellite probe (Cambio, Cambridge, UK).

SNP-based genomic array analysis

For high-resolution detection of genomic imbalances, 300 ng of DNA was extracted using standard methods (DNeasy Blood & Tissue Kit, Qiagen, Valencia, CA) and hybridized to

Illumina Human CNV370-Duo/Quad Genotyping BeadChips (Illumina Inc., San Diego, CA) according to the manufacturer's specifications. Data were first normalized using a proprietary algorithm in the Illumina BeadStudio software (Illumina Inc). Normalized allelic intensity values were thereafter exported and subjected to an additional normalization step using the tQN-software [18]. The tQN software was also used to estimate B-allele frequencies (BAF) for each SNP based on a set of reference genotype clusters. For identification of imbalances, the BAF segmentation software was used, in which BAF-values are transformed into mirrored BAF (mBAF) [19]. BAF segmentation applies an algorithm on the mBAF data to define regions of allelic imbalance. For each resulting segment, a copy number estimate was also given as the median log2 ratio of all SNPs present within the defined segment. Segment copy numbers were estimated using a combination of BAF-segmentation output data and manual inspection of log2 and BAF-plots in comparison to chromosome banding karyotype data. By the automated analysis every consecutive 10 SNPs with mBAF values >0.55 were classified as being in allelic imbalance. To allow detection of minor sub-clones this analysis was complemented by manual inclusion of segments >5 Mb with shifts in mBAF >0.03 compared to the surrounding segments. Segments with log2 ratio >0.073 were classified as genomic gains, those with log2 ratios <-0.080 as genomic losses, and those with log2 ratios between these boundaries as copy number neutral genomic imbalances. Copy number imbalances that were not associated with allelic imbalances (e.g. 4 copies at a 2:2 ratio) were identified by visual inspection of log2 ratio plots in comparison to cytogenetic data. The current detection limit for sub-clonal allelic losses (copy-number reduction and copy-number neutral loss of heterozygosity) and single-copy gains of the HumanCNV370 platform is approximately 20% (17-26%) of analyzed cells at the diploid level, corresponding to 10% of copies for a given segment [19]. Translated to the triploid level of the WiT49 cell lines, this corresponds to a detection limit of approximately 1 of 3 cells (1 of 9 copies).

Statistical analysis

Principal component analysis (PCA) was performed on BAF values using the Qlucore bioinformatics tool (Qlucore AB, Lund, Sweden).

Results and Discussion

In order to assess the potential for formation of clones with novel chromosome alterations in WiT49, we first determined the rate of chromosome segregation errors in WiT49. FISH analysis of dividing WiT49 cells at anaphase stage, using a probe for all human centromeres, demonstrated chromosome segregation errors in 69 (29.5%) of the 235 analyzed cells: chromosome lagging in 10%, chromatin fragment lagging in 3%, anaphase bridging in 5%, and mitotic multipolarity in 11.5% (Fig. 2). Because some chromosome segregation errors, such as sister chromatid non-disjunction, will not be detected by this screening approach [20] the true number of chromosome segregation errors in WiT49 is probably even higher, and it would be fair to assume that WiT49 cells undergo chromosome segregation errors in a least 1/3 of all cell divisions. This rate is high, but nevertheless comparable to other commonly used cancer cell lines [4,8]. In order to assess potential changes in copy number of chromosomal segments resulting from this high level of mitotic missegregation over time, we analyzed WiT49 cells at the beginning (P0) and the end (P10) of a 30-day interval, during which it was subjected 10 times to sub-culturing at a ratio of 1:3 (Fig. 3). In addition, 5 colonies derived from single cells from P0 were expanded in parallel for the same time period in order to simulate the effects of genetic bottlenecks. Continuous monitoring by video microscopy of growing WiT49 cells showed that the average population doubling time of WiT49 cells was approximately 36 h (data not shown), irrespective of colony size, corresponding to approximately 20 population doublings during the period of expansion.

We performed SNP-array analysis of DNA extracted from P0, P10 and the single cell clones M1-M5. At the hypotriploid level of WiT49, the detection limit of sub-clonal genomic imbalances by the HumanCNV370 platform is approximately 1/3 of cells. In a cell population expanding from a single cell clone according to the *clonal liberalism* model with a fitness for cells with novel chromosome changes similar to the other cells in the culture, this implies that any aberrations generated during the first 2 mitoses (generating the first 3 cells) will be detected as sub-clonal abnormalities by SNP-array. Because a total number of five single cell-derived clones were assessed, the probability of detecting sub-clonal aberrations in at least one of the cultures M1-M5 according to this model will be at least $1-(1-f)^{2x5}$, where f is the frequency of chromosome segregation errors. With the minimum frequency of missegregation in WiT49 being 29.5%, the probability of detection sub-clonal aberrations in at least one of the M1-M5 cultures will be at least 98% assuming *clonal liberalism.* Furthermore, this model predicts variability among M1-M5 and also allows for variation between the original WiT49 cell population (P0) and the bulk population analyzed after one month (P10). In contrast, the *clonal conservatism* model with a fitness for cells with novel chromosome aberrations close to 0 predicts a failure to detect sub-clonal aberrations in M1-M5. According to this model the variation between the P0 and P10 will be limited to variations in prevalence of clones present in P0, as clones with novel allelic imbalances are rarely if ever formed in the bulk tumour cell population. However, considering the fact that the P0 population is known to exhibit considerable intercellular cytogenetic heterogeneity, this model nevertheless predicts that there will be genotypic variation among M1-M5 because of the extreme genetic bottleneck provided by single cell cloning.

To analyze the genomic variation among P0, P10, and M1-M5 we first performed PCA of B-allele frequencies for all of the approximately 300,000 polymorphic loci analyzed by the HumanCNV370 array. PCA is a statistical method by which a high number of possibly correlated variables are transformed into a smaller number of uncorrelated variables referred to as principal components, of which the first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible [21]. The B-allele frequency reflects the allelic composition of each polymorphic locus, being 0.5 if the locus is heterozygous with an equal number of A- and B-alleles (AB, AABB *etc.*), 1 if the locus is homozygous for B (B, BB, BBB *etc.*), and 0 if the locus is homozygous for A (A, AA, AAA etc). Depending on the overall number of A and B alleles in the analyzed DNA, the B-allele frequency may thus vary between 1 and 0 and cell populations with similar genotypes will have B-allele frequencies similar to each other. However, B-allele frequencies cannot distinguish cell populations that differ from each other through whole genome duplication. Chromosome analysis of P0, P10 and M1-M5 showed a modal number of 60-65 for all of the lines, making their B-allele frequencies comparable. The results of the PCA showed a close similarity in genotype between P10 and P0, while M1-M5 exhibited a high degree of genotype diversity compared to P0/P10 (Fig. 4A). The PCA also revealed clear similarities between M2 and M5, on the one hand, and some similarity among M1, M3 and M4, on the other hand. This pattern was maintained when different levels of filtration were applied to the data set, including 30-100% of SNPs. Thus a larger degree of variation was present among the single cell-derived sub-lines and between these lines and P0/P10, than between P0 and P10.

To identify differences in genotype between the cell lines, we first performed PCA of the SNP loci, in which the loci showing a high degree of variation are expected to separate according

to the principal components (Fig. 4B). Furthermore, individual SNPs that are located close to each other on the same chromosome will co-vary and form clusters in the PCA plot because they are likely to be present at an equal B-allele frequency. Each such chromosome segment will be represented by two equally sized clusters at diametrically opposite locations in the PCA plot. Three distinct clusters were observed at each edge of the PCA plot. These clusters corresponded to SNPs located in three contiguous chromosomal segments, *i.e.* 1p36-q32, 12q14-q23, and 13q12-q31, indicating that these three segments were responsible for a major part of the variation among P0, P10 and M1-M5. To validate this finding, we then performed combined manual and automated analysis of B-allele frequency and log2 data on chromosomes 1, 12 and 13 for all the sub-lines. We thereby determined the copy number of each chromosome segment, including assessment of sub-clonal variation, *i.e.* the presence of intercellular heterogeneity with respect to copy number in each line (Fig. 5). This revealed that the three segments identified as highly variable by the PCA indeed showed a high degree of variation among the seven cell lines. In all these segments, there was sub-clonal copy number variation present in P0 and P10, but not in M1-M5. There was also a high degree of similarity between M1, M3 and M4 on the one hand, and M2 and M5 on the other hand, in accordance to the results by PCA. M2 and M5 showed completely similar copy numbers for the three segments, while M1, M3 and M4 were identical with respect to 1p36-q32 and 12q14-q23 but showed some variation in 13q12-q31. However, variation was also present in minor chromosome areas outside the segments identified as hypervariable by PCA, indicating that this approach was not sufficient for determining the full spectrum of genomic variation among the sublines. Furthermore, the PCA approach favored identification of large variable genomic segments while discriminating minor segments that may have an equally high degree of variation.

We therefore proceeded with combined automated and manual detection of genomic copy number variation among the lines (Figure 5; Supplemental Fig. 1). This revealed sub-clonal copy number variation in P0, P10, or both for the segments 1p36-q44, 5p14, 6q14-q21, 7q31-q36, 9q11-q22, 12q14-q24, 13q12-q21, 13q31, 16p12-q12, 16q23, 19p13, 21q21, and 21q22 (Table 1). The sub-clonal copy number variants in P10, were found also in the P0 main clone, in P0 sub-clones or in two or more of M1-M5, indicating that they were present already before the first sub-culture separating P0 and P10. Besides this sub-clonal variation, there was little difference between P0 and P10. In fact, only two segments showed a different copy number in the main clone of one of these lines but not in the other: 9q11-q22 and 6q14-q21. However, in 6q14-q21 the copy number variant found in P10 but not in P0 was also present in M2 and M5, indicating its presence also in P0, from which M2 and M5 were initially derived, albeit in cell populations too small for detection by SNP-array. Thus, only one segmental chromosome imbalance potentially acquired during the 1 month in culture from P0 and P10 could be detected. There was a considerably higher diversity among M1-M5, well in accordance with the PCA results (Table 2). On average, M clones showed copy number alterations compared to P0 in 5.8 chromosome segments (range 3-10), and when compared to each other in an average of 8.9 segments (range 2-16; Figure 6; Supplemental Tables 1 and 2). In total, 23 continuous chromosome segments exhibited variation in copy number. Of these, 10 exhibited variants that were part of the spectrum of sub-clonal variation in P0, P10 or both, indicating that they had not been acquired after single-cell cloning. Four segments were found in more than one of the lines M1-M5, implying that they also were present before single-cell cloning took place, but in cell populations that were too small for detection at analysis of P0 or P10 but yet present in more than one of the single cells from which M1-M5 originated. Nine continuous segments showed copy number variants that were present in single sub-lines. There were no sub-clonal copy number variants detected in any of M1-M5. Hence, no

chromosomal imbalance that with certainty had formed during the first 2 mitoses after seeding each of the five single cells giving rise to M1-M5 could be detected. Considering that the probability of detecting at least one such abnormality by our SNP-array platform was 98% if cells with novel genomic imbalances have a fitness equal to that of other cells, these negative results argue against the *clonal liberalism* model. Instead, the findings indicate that the genetic bottleneck introduced by single cell cloning was responsible for the main part, if not all of the variation among M1-M5, supporting *clonal conservatism*.

The significance of the copy number variants present uniquely in single sublines is equivocal. Because none of the nine segments with unique copy numbers in M1-M5 showed evidence of being present in <100% of the cells, the most likely explanation is that they were present already in the single cell giving rise to each clone, thus also being products of the bottle neck effect. On the other hand, it cannot be excluded that selection at the beginning of clone formation resulted in overgrowth of cell populations having acquired novel unique aberrations formed after explantation of single cells from P0. Evaluation of the presence or absence of the unique copy number variants in P0 could potentially differentiate between these two possibilities. However, for most variable segments this would require the detection of small sub-clones in P0 by single copy FISH probes. The prevalence of such clones in P0 would be expected to be low as they were not detected by the SNP-array and close to the background levels of single copy probe FISH. The resulting data would therefore often be difficult to validate. The largest unique copy number variant in M1-M5 was in 2pter-2qter, for which M5 showed a loss of an entire copy of chromosome 2. This copy number variant was thus potentially traceable in P0 without the need for FISH detection of small genomic segments. Chromosome banding analysis complemented by metaphase FISH showed that the stem line of P0 contained three ostensibly normal copies of chromosome 2 and one copy with loss of

2q32-qter through an unbalanced translocation with chromosome 8 (Supplemental Fig. 2). However, in 17% of the analyzed P0 cells (3/18), below the detection level of the SNP-array platform, there was a loss of a chromosome 2 homologue, corresponding to the finding in M5. Hence, the cytogenetic data provided support for the origin of this aberration from a sub-clone in P0 having passed through the genetic bottle neck provided by single cell cloning, similar to the majority of other copy number variants differentiating M1-M5.

Conclusion

In summary, we evaluated the *in vitro* genome evolution of a cancer cell line with a high rate of mitotic segregation anomalies. In particular, we tested whether the proliferative survival of cells with novel chromosome aberrations was sufficiently high (*clonal liberalism*) to produce changes in the cell line's genotype during prolonged growth (1 month, 10 passages, 20 populations doublings). This was compared to a model with low fitness of cells with novel aberrations (*clonal conservatism*) according to which a change in a cell line's genotype can occur readily by shifting the relative balance of pre-existing sub-clones through a genetic bottleneck but according to which clones with novel chromosome aberrations are rarely if ever formed. Presuming equal fitness of cells having acquired novel chromosome aberrations compared to surrounding competing cells according to *clonal liberalism*, our experimental system should have detected at least one sub-clonal novel aberration in the single cell-derived clones M1-M5. Furthermore, this model would allow for differences between P0 and P10 as minor clones expanded on equal terms as those already present in P0. Neither of these two predictions was consistent with our experimental data. Instead, we found little difference in genotype between P0 and P10 while the population bottlenecks through which M1-M5 were formed resulted in a high degree of genomic diversity in which most of the copy number variants differentiating the M-lines from each other could be traced back to pre-existing sub-

clones in the original bulk cell line population, well in accordance to *clonal conservatism*. These findings show that even a cell line with a very high rate of mitotic instability can largely maintain its genotype if cultured as a bulk population, at least during periods limited to a few weeks. The differences that do occur are largely caused by subtle shifts in the prevalence of pre-existing clones. Nevertheless, genetic bottlenecks created by any event that reduces the proliferating population to a minimum can result in dramatic shifts in genotype as sub-clones too small for being detected by genetic screening of the original bulk population can be selected for expansion. This should be an important observation as a many manipulations of cells for functional studies are associated with a risk of increased cell death, proliferative arrest of majority cell population, or other types of clonal selection that could introduce genetic bottlenecks and thereby a shift in genotype. In such experiments, the post-manipulation cell population is typically compared experimentally with undisturbed cells from the same line and/or cells subjected to a mock-manipulation that could introduce a genetic bottleneck on its own. The present study shows that such comparisons may not always be warranted, as the cells used for comparison may not only differ in the respect meant to be tested (expression of a candidate gene, exposure to a biologically active substance *etc*.) but may also have differences in genotype that could bias data interpretation. Finally, the experiments detailed in this paper may also have some correlation to the clinical situation of recurrent tumour disease. Sub-optimal treatment resulting in a small population of surviving tumour cells can result in an *in vivo* genetic bottleneck that could explain major changes in genotype between primary tumours and recurrent tumours. One could also envision a similar genetic bottleneck occurring during the metastatic process. In this context, it should be stressed that the present study is limited to the *in vitro* situation, where both mutation rate and micro-environmental selection pressure could differ significantly from *in vivo* conditions. Our study therefore prompts further investigation of the potential effects of genetic bottle necks *in*

vivo, for example in animal models of tumour disease and in human-animal tumour xenografts. Compared to the vast resources spent in examining other causes of genomic variability in cancer cells, the role played by genetic bottlenecks has been surprisingly little explored. The present study suggests that this phenomenon is worthy of more attention than previously assumed.

Acknowledgments

This study was supported by the Swedish Childhood Cancer Foundation, the Swedish Cancer Society, the Swedish National Research Council, the Lund University Hospital Donation Funds, the Gunnar Nilsson Cancer Foundation, the Crafoord Foundation, the Erik-Philip Sörensen Foundation, the Lundgren Foundation, the Schyberg Foundation, and the Medical Faculty at Lund University.

References

[1] M. Lacroix, G. Leclercq, Relevance of breast cancer cell lines as models for breast tumours: an update, Breast Cancer Res. Treat. 83 (2004) 249-289.

[2] C. Horrocks, R. Halse, R. Suzuki, P.R. Shepherd, Human cell systems for drug discovery, Curr. Opin. Drug Discov. Devel. 6 (2003) 570-575.

[3] S.F. Stinson, M.C. Alley, W.C. Kopp, H.H. Fiebig, L.A. Mullendore, A.F. Pittman, S. Kenney, J. Keller, M.R. Boyd, Morphological and immunocytochemical characteristics of human tumor cell lines for use in a disease-oriented anticancer drug screen, Anticancer Res. 12 (1992) 1035-1053.

[4] C. Lengauer, K.W. Kinzler, B. Vogelstein, Genetic instability in colorectal cancers, Nature 386 (1997) 623-627.

[5] B.M. Ghadimi, D.L. Sackett, M.J. Difilippantonio, E. Schröck, T. Neumann, A. Jauho, G. Auer, T. Ried, Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations, Genes Chromosomes Cancer 27 (2000) 183-190.

[6] D. Gisselsson, L. Pettersson, M. Höglund, M. Heidenblad, L. Gorunova, J. Wiegant, F. Mertens, P. Dal Cin, F. Mitelman, N. Mandahl, Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 5357-5362.

[7] Y. Stewénius, L. Gorunova, T. Jonson, N. Larsson, M. Höglund, N. Mandahl, F. Mertens, F. Mitelman, D. Gisselsson, Structural and numerical chromosome changes in colon cancer develop through telomere-mediated anaphase bridges, not through mitotic multipolarity, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 5541-5546.

[8] N.J. Ganem, S.A. Godinho, D. Pellman, A mechanism linking extra centrosomes to chromosomal instability, Nature 460 (2009) 278-282.

[9] S.E. Mamaeva, Karyotypic evolution of cells in culture: a new concept, Int. Rev. Cytol. 178 (1998) 1-40.

[10] M.L. Larramendy, T. Lushnikova, A.M. Bjorkqvist, Wistuba, II, A.K. Virmani, N. Shivapurkar, A.F. Gazdar, S. Knuutila, Comparative genomic hybridization reveals complex genetic changes in primary breast cancer tumors and their cell lines, Cancer Genet. Cytogenet. 119 (2000) 132-138.

[11] T. Knutsen, H.M. Padilla-Nash, D. Wangsa, L. Barenboim-Stapleton, J. Camps, N. McNeil, M.J. Difilippantonio, T. Ried, Definitive molecular cytogenetic characterization of 15 colorectal cancer cell lines, Genes Chromosomes Cancer 49 (2009) 204-223.

[12] M. Nugoli, P. Chuchana, J. Vendrell, B. Orsetti, L. Ursule, C. Nguyen, D. Birnbaum, E.J. Douzery, P. Cohen, C. Theillet, Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications, BMC Cancer 3 (2003) 13.

[13] J. Alami, B.R. Williams, H. Yeger, Derivation and characterization of a Wilms' tumour cell line, WiT 49, Int. J. Cancer 107 (2003) 365-374.

[14] Y. Stewénius, Y. Jin, I. Øra, A. Frigyesi, J. Alumets, B. Sandstedt, H. Bras, J. De Kraker, F. Mertens, D. Gisselsson, Defective chromosome segregation and telomere dysfunction in aggressive Wilms tumours Clin. Cancer Res. 13 (2007) 6593-6602.

[15] Y. Stewenius, Y. Jin, I. Ora, I. Panagopoulos, E. Moller, F. Mertens, B. Sandstedt, J. Alumets, M. Akerman, J.H. Merks, J. de Kraker, D. Gisselsson, High-resolution molecular cytogenetic analysis of Wilms tumors highlights diagnostic difficulties among small round cell kidney tumors, Genes Chromosomes Cancer 47 (2008) 845-852.

[16] D. Gisselsson, Refined characterisation of chromosome aberrations in tumours by multicolour banding and electronic mapping resources, Methods Cell Sci. 23 (2001) 23-28.

[17] D. Gisselsson, M. Lv, S.W. Tsao, C. Man, C. Jin, M. Hoglund, Y.L. Kwong, Y. Jin, Telomere-mediated mitotic disturbances in immortalized ovarian epithelial cells reproduce chromosomal losses and breakpoints from ovarian carcinoma, Genes Chromosomes Cancer 42 (2005) 22-33.

[18] J. Staaf, J. Vallon-Christersson, D. Lindgren, G. Juliusson, R. Rosenquist, M. Hoglund, A. Borg, M. Ringner, Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios, BMC Bioinformatics 9 (2008) 409.

[19] J. Staaf, D. Lindgren, J. Vallon-Christersson, A. Isaksson, H. Goransson, G. Juliusson, R. Rosenquist, M. Hoglund, A. Borg, M. Ringner, Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays, Genome Biol. 9 (2008) R136.

[20] D. Gisselsson, Classification of chromosome segregation errors in cancer, Chromosoma 117 (2008) 511-519.

[21] I. Jolliffe, Principal Component Analysis, 2, Springer-Verlag New York Inc. , New York, 2002.

Figure legends

Fig. 1. Models for genome evolution. Contrasting evolutionary models for cell populations with a high formation rate of novel chromosome aberrations. *Clonal liberalism* (top), assumes that cells having acquired novel chromosome aberrations (denoted by colour shifts) will have a fitness comparable to other cells and may give rise to cell populations with novel chromosome changes during growth of the bulk cell line population from passage 0 (P0) to passage 10 (P10). Under the provided experimental conditions, expansion of single cells (M) working according to this model will result in detection of at least one sub-clone with novel chromosome aberrations at a probability close to 100%. *Clonal conservatism* (bottom), assumes that the fitness of cells having acquired novel chromosome aberrations is too low to allow the emergence of cell populations with novel chromosome changes during growth of the bulk cell line population (dead/non-proliferative cells denoted by crosses). Variation in genotype during growth of the bulk cell line population can thus result only from changes in the relative prevalence of pre-existing sub-clones. Expansion of single cells working according to this model will result in cell populations without detectable genetically different sub-clones and with variation compared to the original cell population emerging only as a result of the genetic bottleneck provided by single cell cloning. Thus, the genomic profiles of M populations should always be traceable to variations present at P0.

Fig. 2. Mitotic segregation errors. WiT49 cells at anaphase hybridized with centromeric probes covering all chromosomes. (A) Bipolar anaphase cells with a single lagging chromosome (arrow). (B) Tripolar anaphase cells with multiple lagging chromosomes (arrows).

Fig. 3. Cell culture systems. Five single WiT49 cells were isolated from the original bulk population (P0) and allowed to expand for approximately 20 population doublings (one month), forming the single cell-derived populations M1-M5. In parallel, P0 was sub-cultured at a 3:1 ratio 10 times to form the secondary bulk population P10.

Fig. 4. Data mining of SNP-array profiles. (A) Principal component analysis (PCA) including the 90,000 polymorphic with the highest degree of variability in the dataset shows a clear sub-division of M-clones in groups consisting of M1, M3, M4, and M2, M5, respectively. P0 and P10 show a closer similarity to each other than to any of the clones M1-M5. Lines indicate nearest-neighbours in the PCA matrix. (B) PCA of individual SNP-loci visualise large chromosome segments accounting for a high degree of variability in the dataset. Clusters of SNPs mapping to the same chromosome segment are highlighted in red, blue and green.

Fig. 5. Segment copy numbers. For each segment, the chromosome homologue with the highest copy number is drawn to the left. The copy number estimate for each chromosome segment (colour key in panel to the right) was based on allelic and log2 ratios obtained by SNP-array combined with cytogenetic data. Single cell-derived clones (M1-M5) with unique copy number variants are denoted by red letters, while lines with variants present also in at least one other M-line are denoted by green letters. In P0 and P10, sub-clonal copy number variants are drawn as parallel lines outside the continuous lines indicating majority copy numbers. The segments in chromosomes 1, 12 and 13 identified as highly variable by principal component analysis are coloured (see Fig. 4). Segments with unique copy number variants in M1-M5 are indicated by red brackets, segments with variants present in more than one M-line by orange brackets, and segments with variants present also in the bulk

populations (P0 and/or P10) by blue brackets. The full copy number analysis is shown in Supplemental Fig. 1.

Fig. 6. Variation among sub-lines. Line colour reflects the number of chromosome segments showing copy number variability between specific lines (colour key to the right). The figure is based on data on the exact number of variable segments provided in Supplemental Tables 1 and 2.

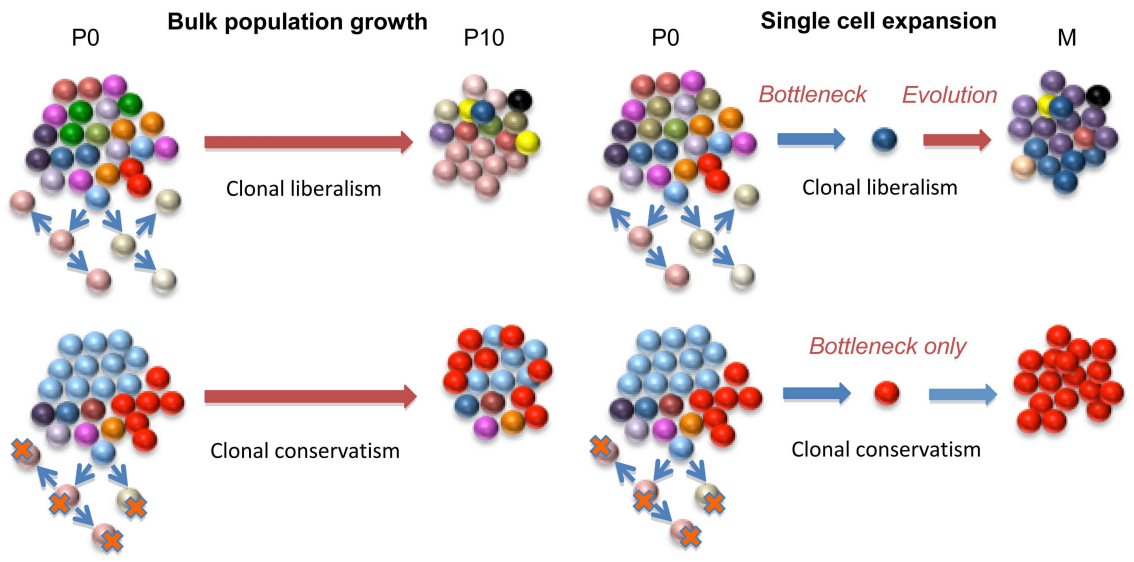
Supplemental Material

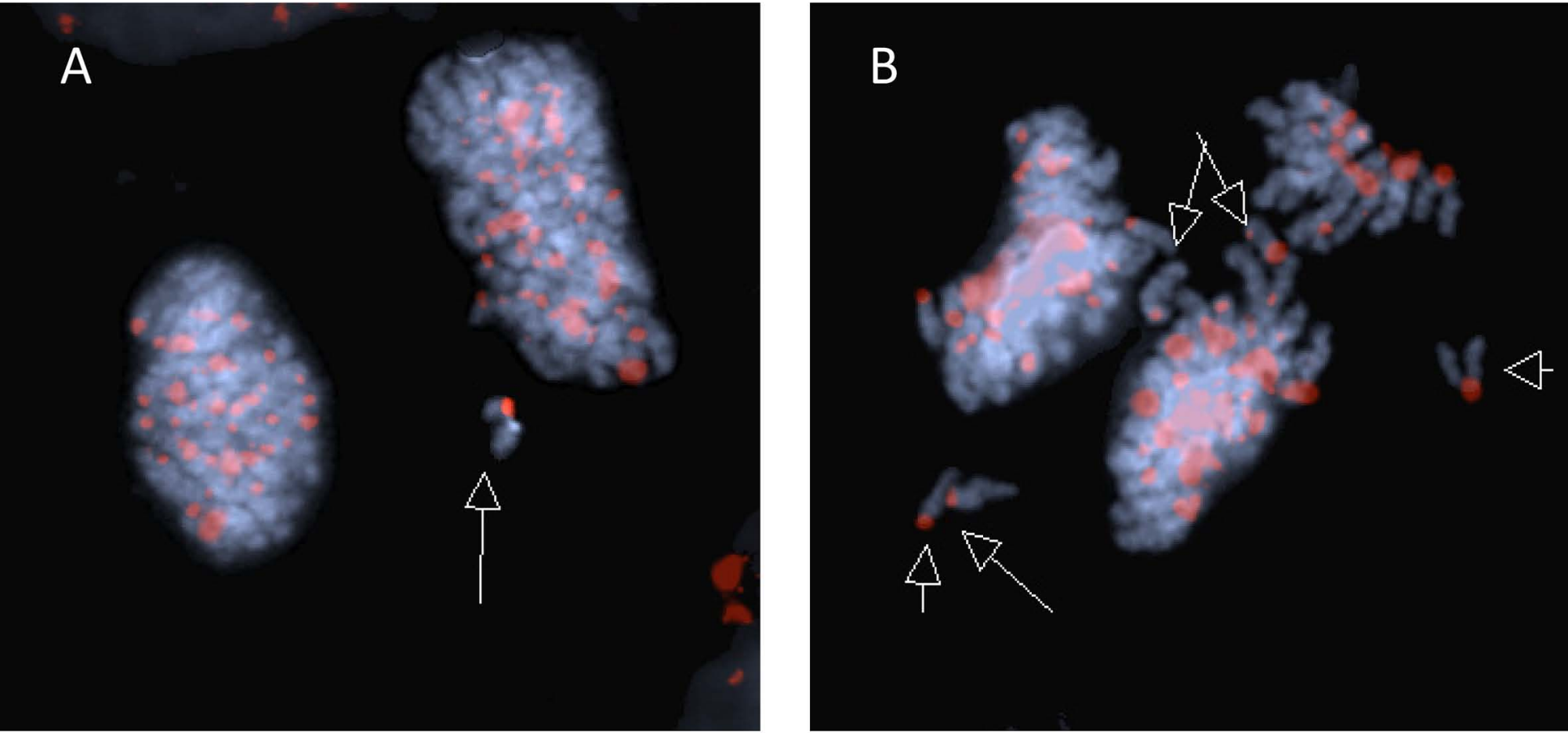
Supplemental Fig. 1. Segment copy numbers for all chromosomes. The copy number estimate for each chromosome segment is based on allelic and log2 ratios obtained by SNP-array combined with cytogenetic data. For each segment, the chromosome homologue with the highest copy number is drawn to the left. See legend to Fig. 5 for annotations.

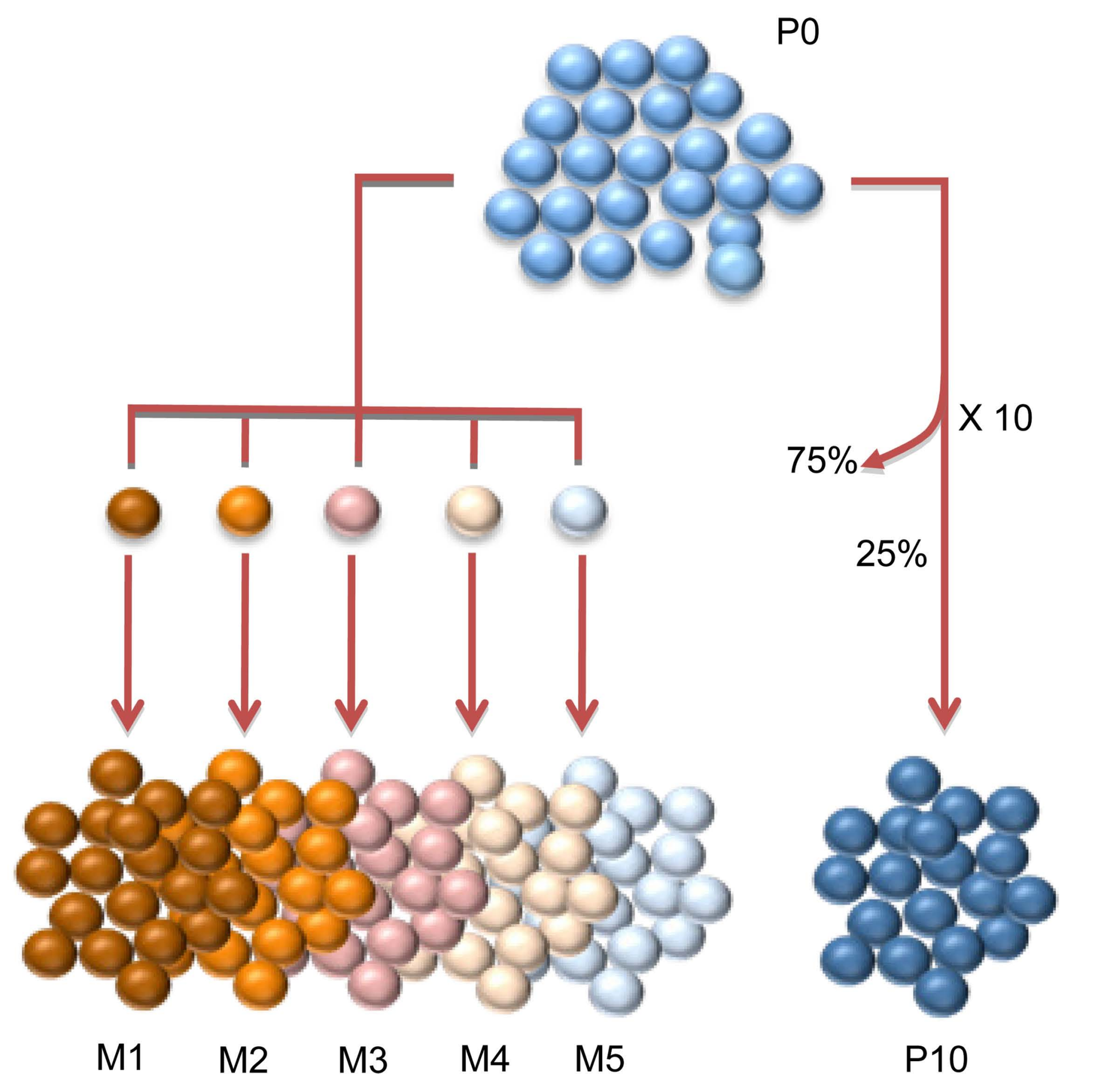
Supplemental Fig. 2. Metaphase fluorescence in situ hybridization using whole chromosome painting probes for chromosomes 2 (green), 8 (blue), and 4 (red) shows an unbalanced 2;8 translocation (arrow).

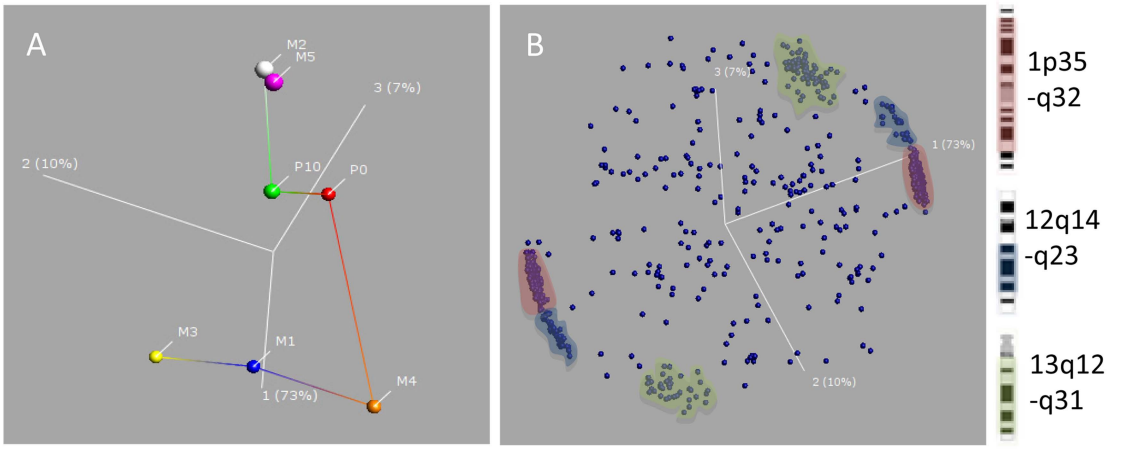
Supplemental Table 1. Segments with variation in copy number between specific lines.

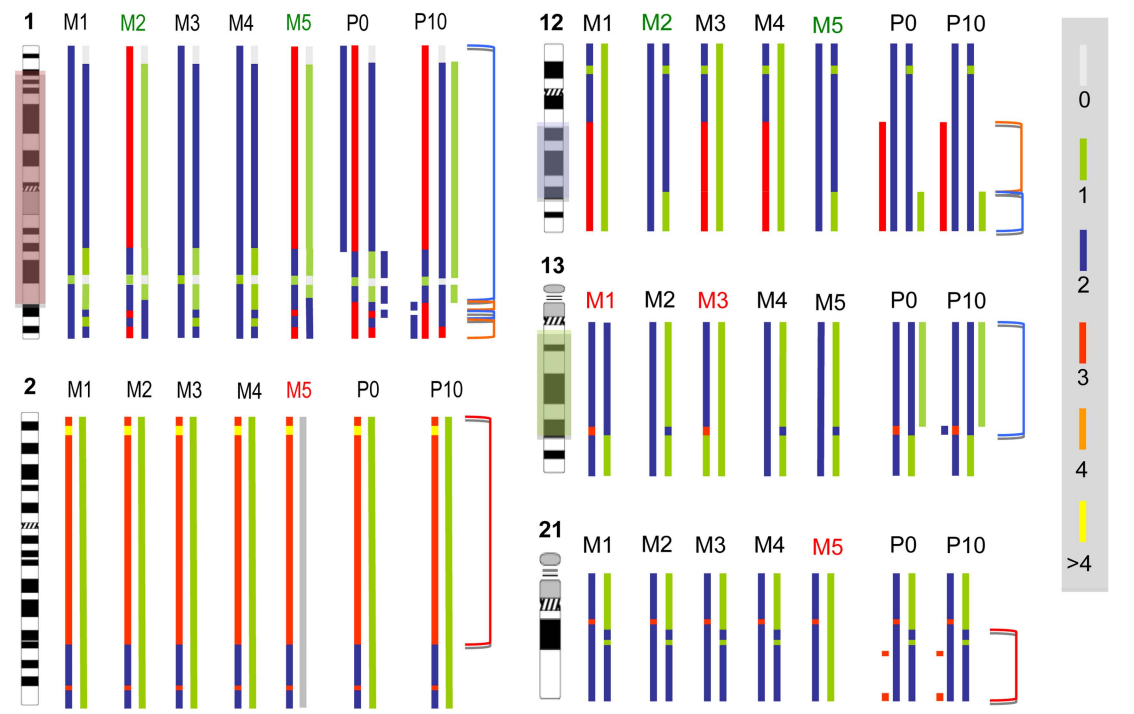
Supplemental Table 2. Number of segments exhibiting variation between specific lines.


Bulk population growth
P0
P10
Clonal liberalism
Clonal conservatism
Single cell expansion
P0
M
Bottleneck
Evolution
Clonal liberalism
Bottleneck only
Clonal conservatism


A
B


P0
X 10
75%
25%
M1
M2
M3
M4
M5
P10


A
M2
M5
3 (7%)
2 (10%)
P10
P0
M3
M1
M4
1 (73%)
B
3 (7%)
1 (73%)
2 (10%)
1p35
-q32
12q14
-q23
13q12
-q31


1
M1
M2
M3
M4
M5
P0
P10
2
M1
M2
M3
M4
M5
P0
P10
12
M1
M2
M3
M4
M5
P0
P10
13
M1
M2
M3
M4
M5
P0
P10
21
M1
M2
M3
M4
M5
P0
P10
0
1
2
3
4
>4

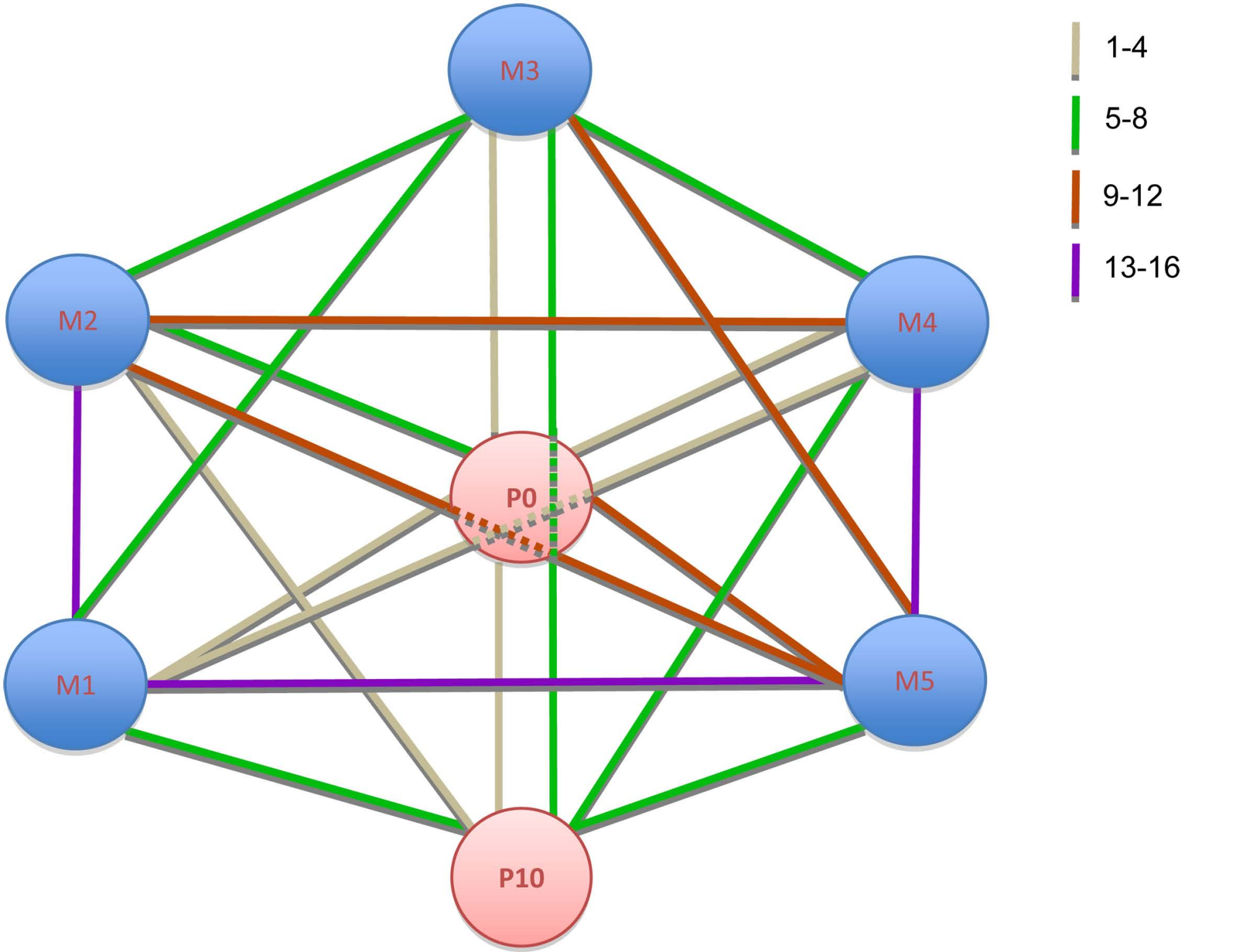

M3
M2
M4
P0
M1
M5
P10
1-4
5-8
9-12
13-16

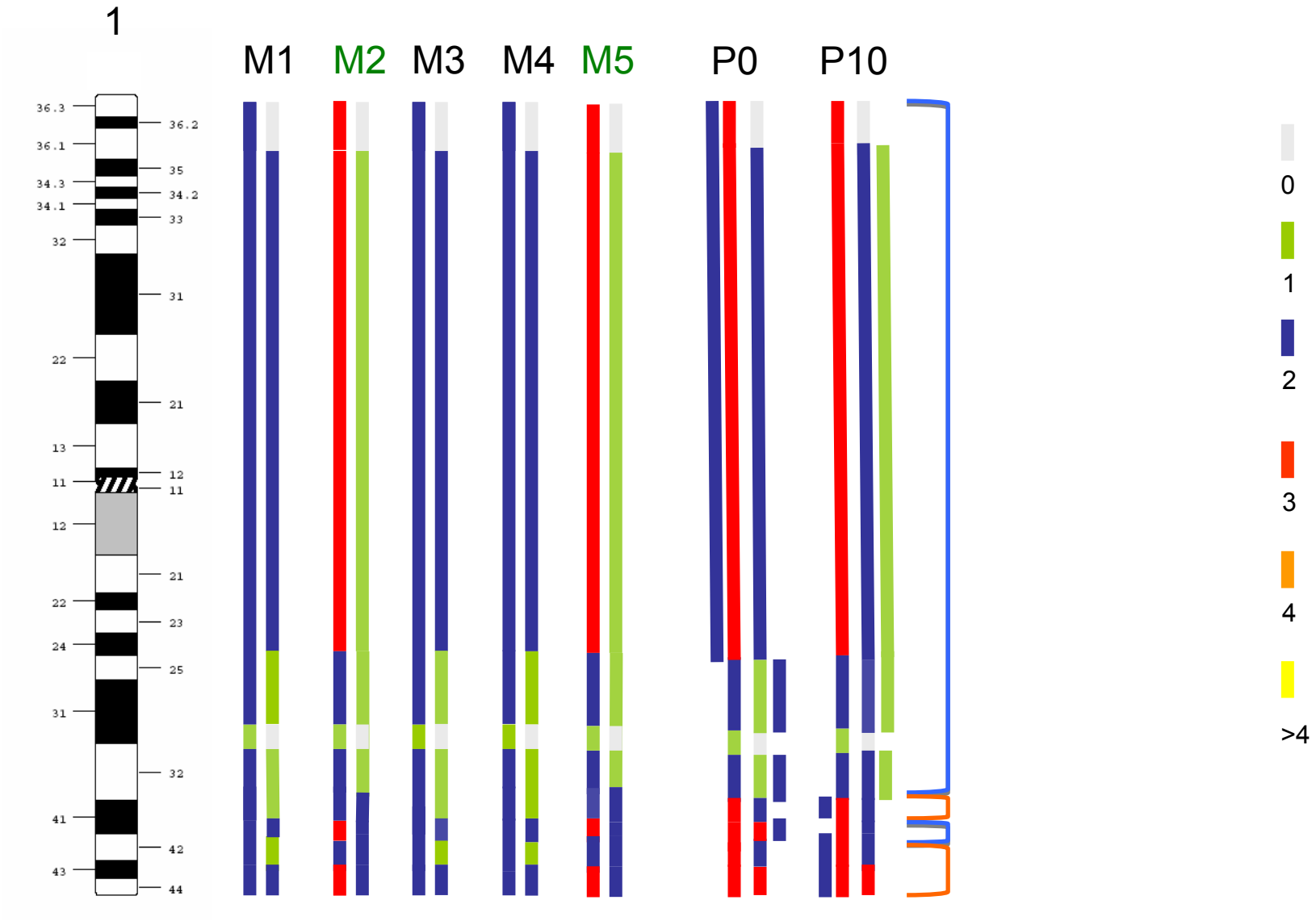
Table 1. Copy number variation between P0 and P10[a]

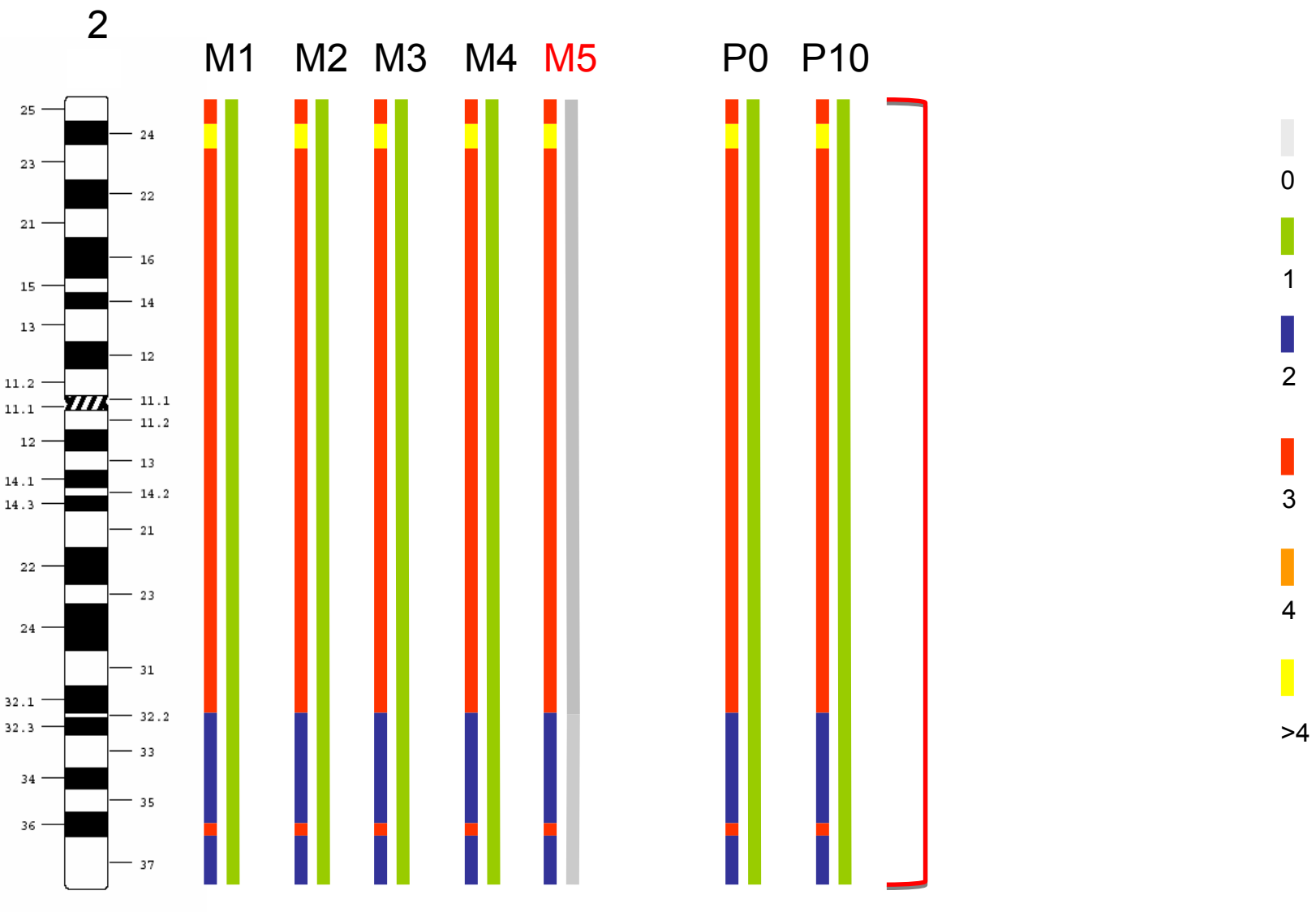
Alteration	Variable segments	Number of segments
Shifts between main clone and sub-clone in P0/P10	1q35-q32, 1q41-q42, 6q14-q21, 6q21	4
Present in P10 sub-clone and any of M1-M5 but not in P0	1p35-q25, 1q41, 1q42, 1q43-qter, 13q31	5
Present in P10 main clone and any of M1-M5 but not in P0	6q14-q21	1
Present only in P10	9q11-q22	1

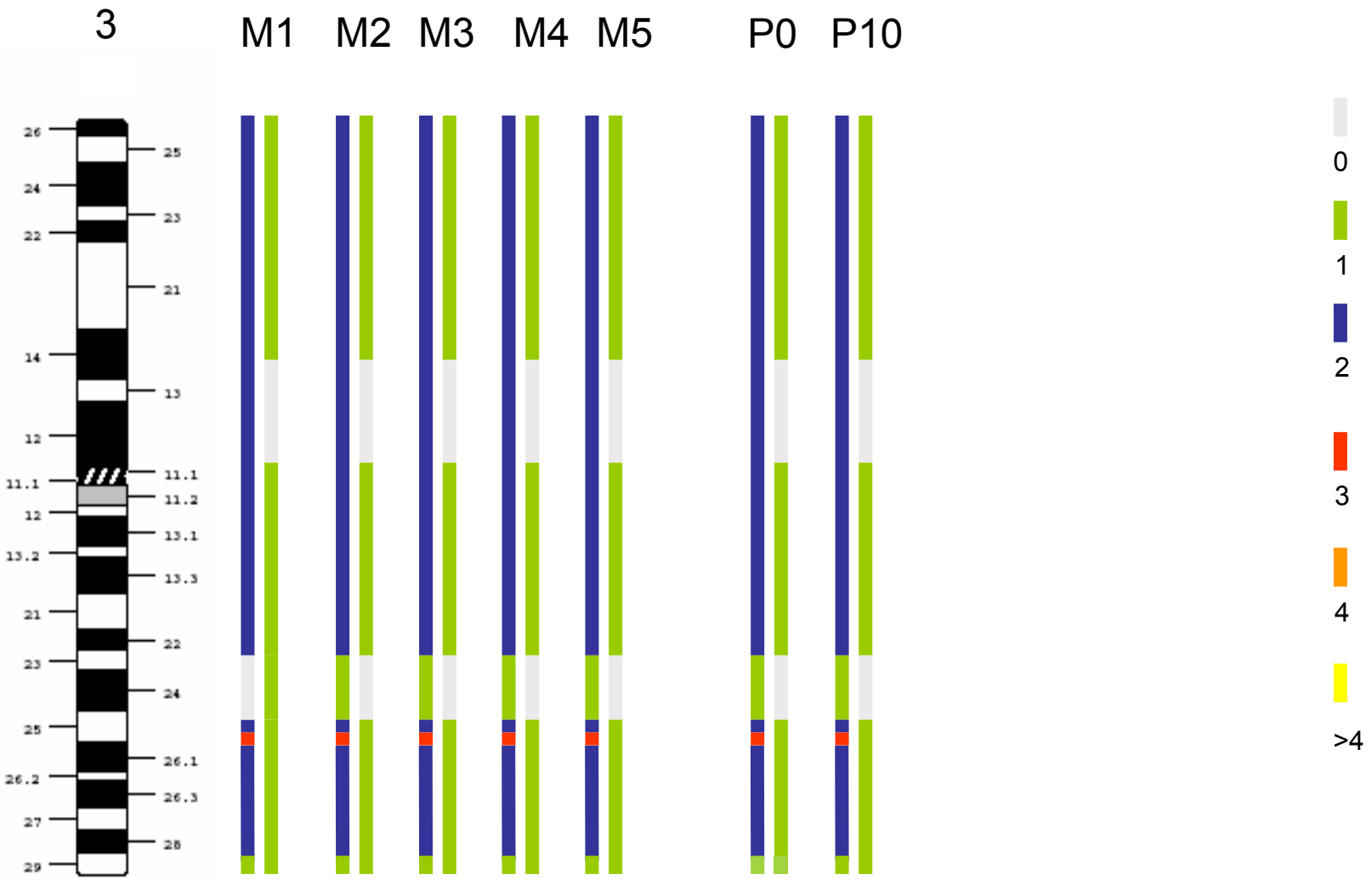
[a] Including main clone and sub-clones in P10 and assuming that copy number variants found in M1-M5 and P10 but not in P0 correspond sub-clones below detection limit in P0.

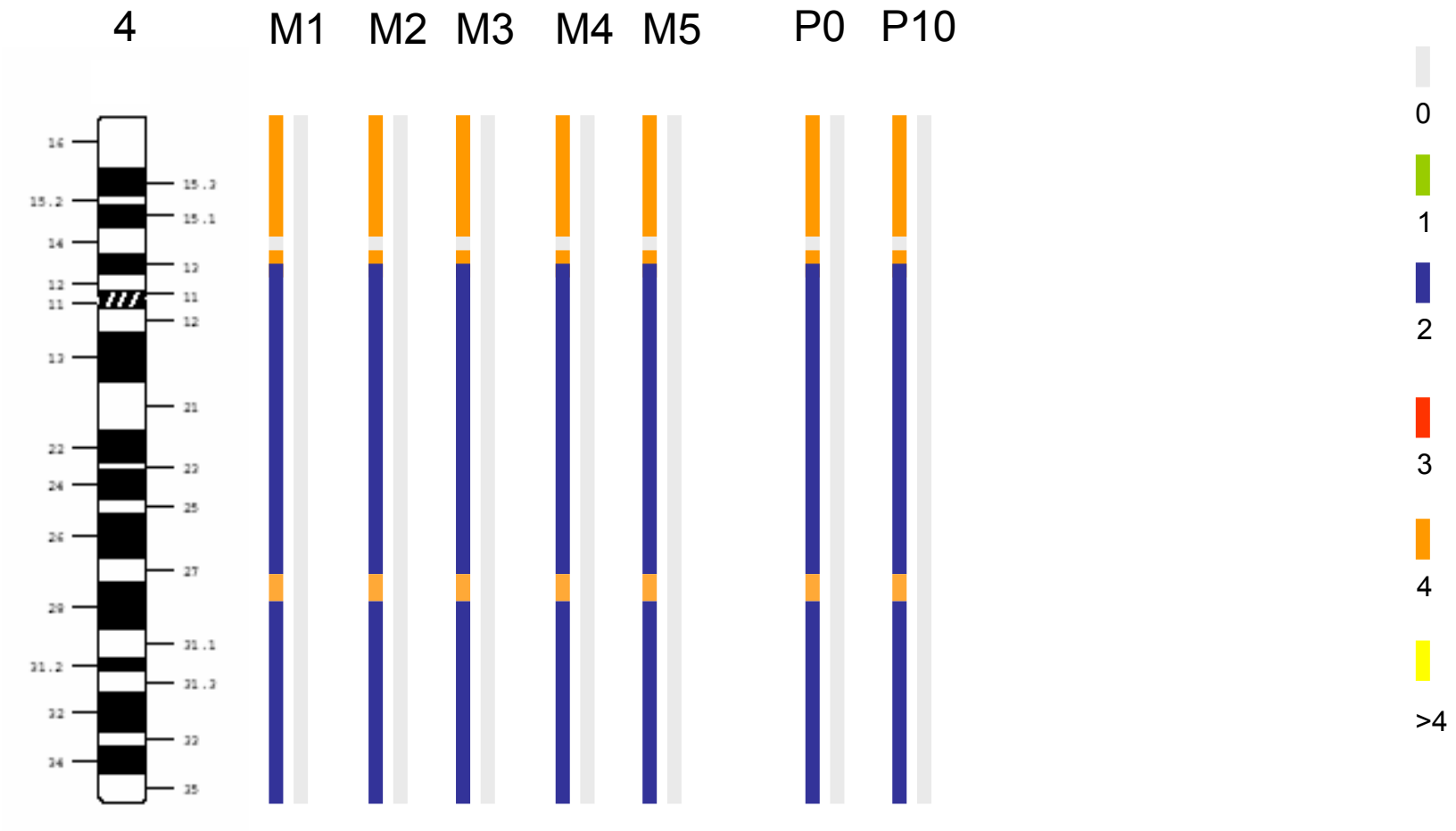
Table 2. Copy number variation among M1-M5

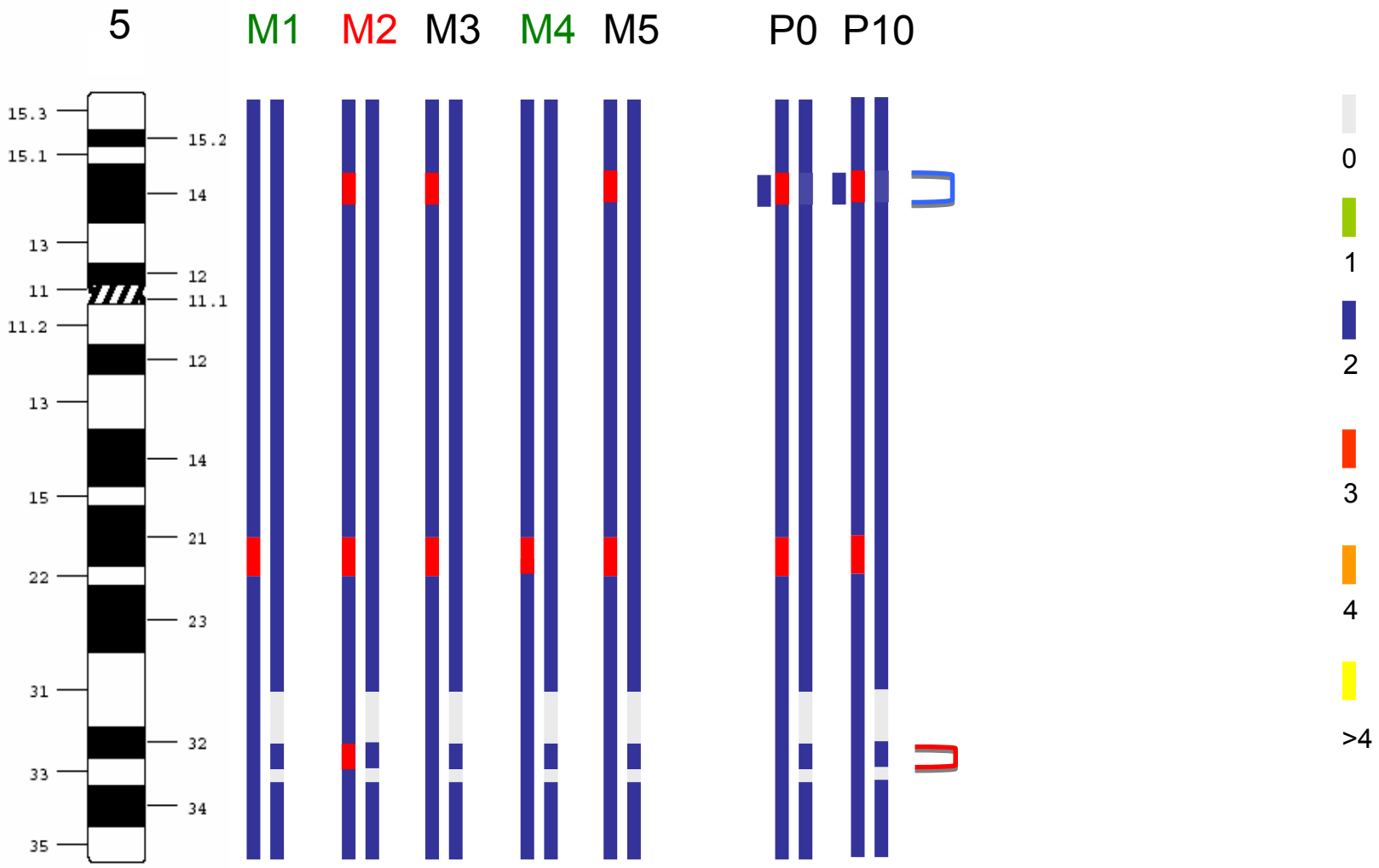
Alteration	Variable segments	Number of segments
Present in P0/10	1pter-q32, 1q41, 5p14, 6q14-q21, 7q32-qter, 12q23-qter, 13q11-q31, 16p12-q12, 16q23, 19p13	10
Present in >1 of M1-M5	1q32-q41, 1q42-qter, 12q13-q23, 17q11	4
Present only in one of M1-M5	2pter-2qter (M5), 5q32 (M2), 9q32-q32 (M5), 10q22 (M3), 11p15-p11 (M2), 10p12-q24 (M5), 15q22-qter (M5), 19q11-q13 (M1), 21q11-21qter (M5)	9
Sub-clonal	-	0

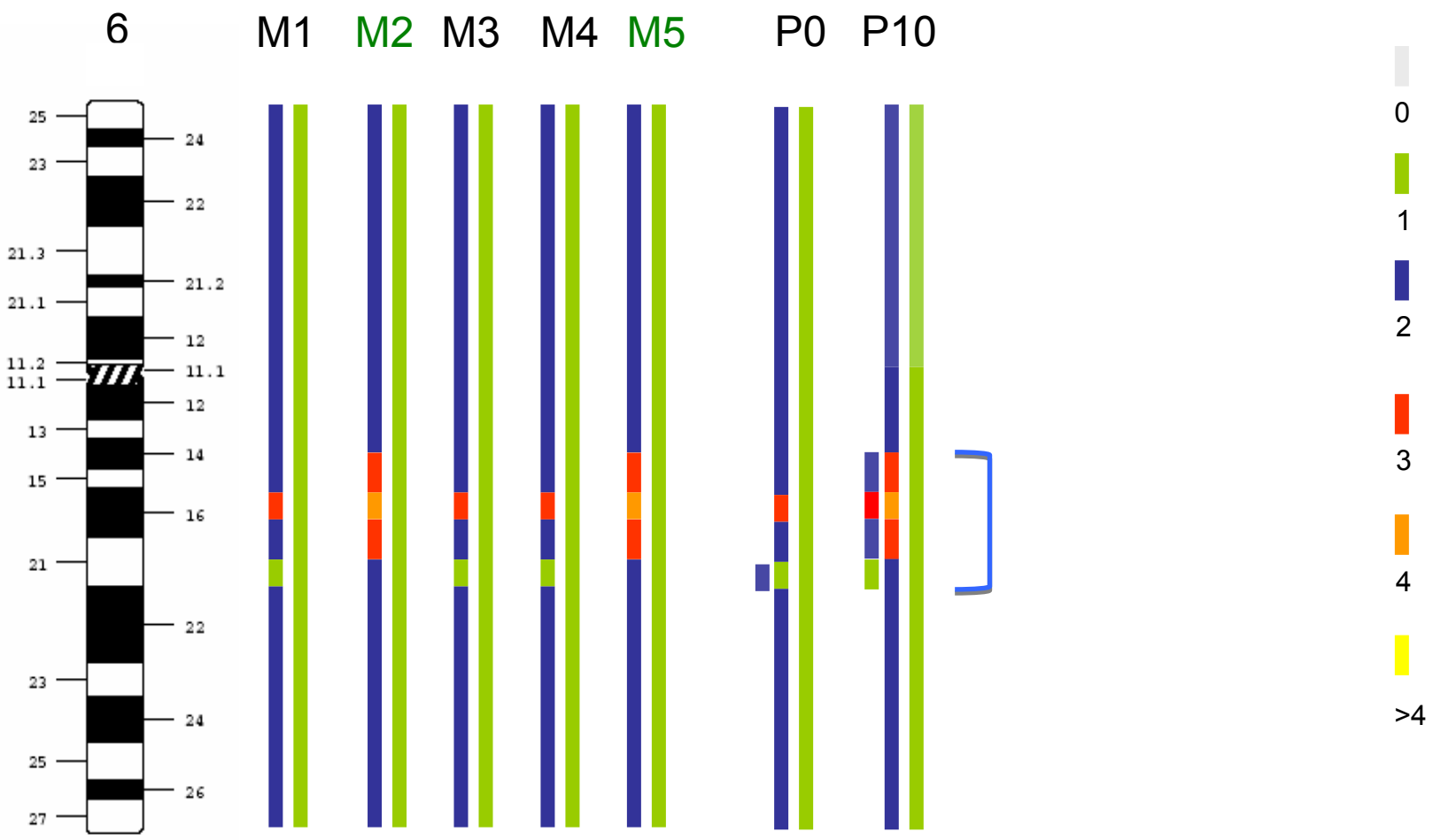

1
M1
M2
M3
M4
M5
P0
P10
36.3
36.2
36.1
35
34.3
34.2
34.1
33
32
31
22
21
13
12
11
11
12
21
22
23
24
25
31
32
41
42
43
44
0
1
2
3
4
>4

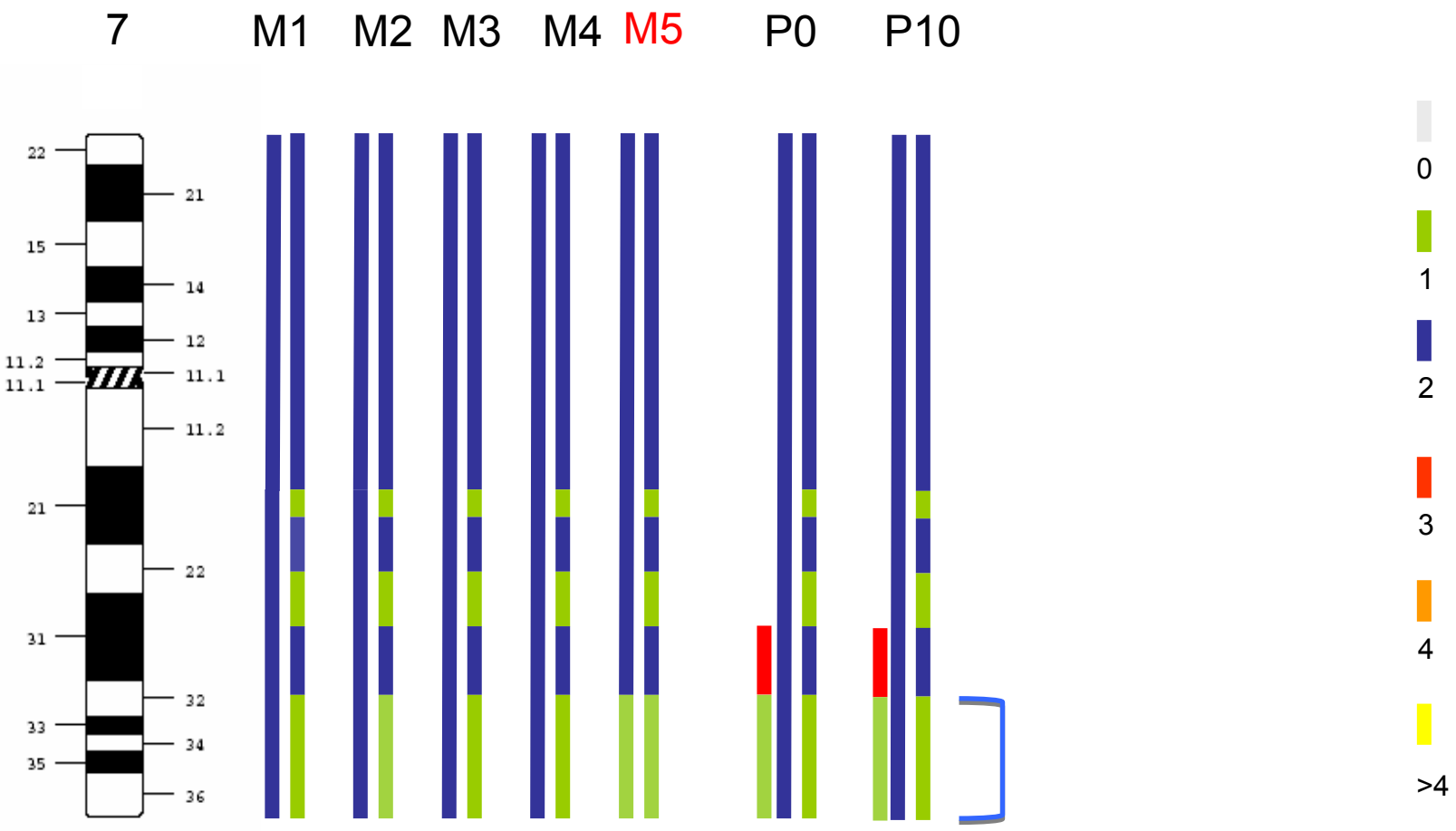

2
M1
M2
M3
M4
M5
P0
P10
25
24
23
22
21
16
15
14
13
12
11.2
11.1
11.1
11.2
12
13
14.1
14.2
14.3
21
22
23
24
31
32.1
32.2
32.3
33
34
35
36
37
0
1
2
3
4
>4

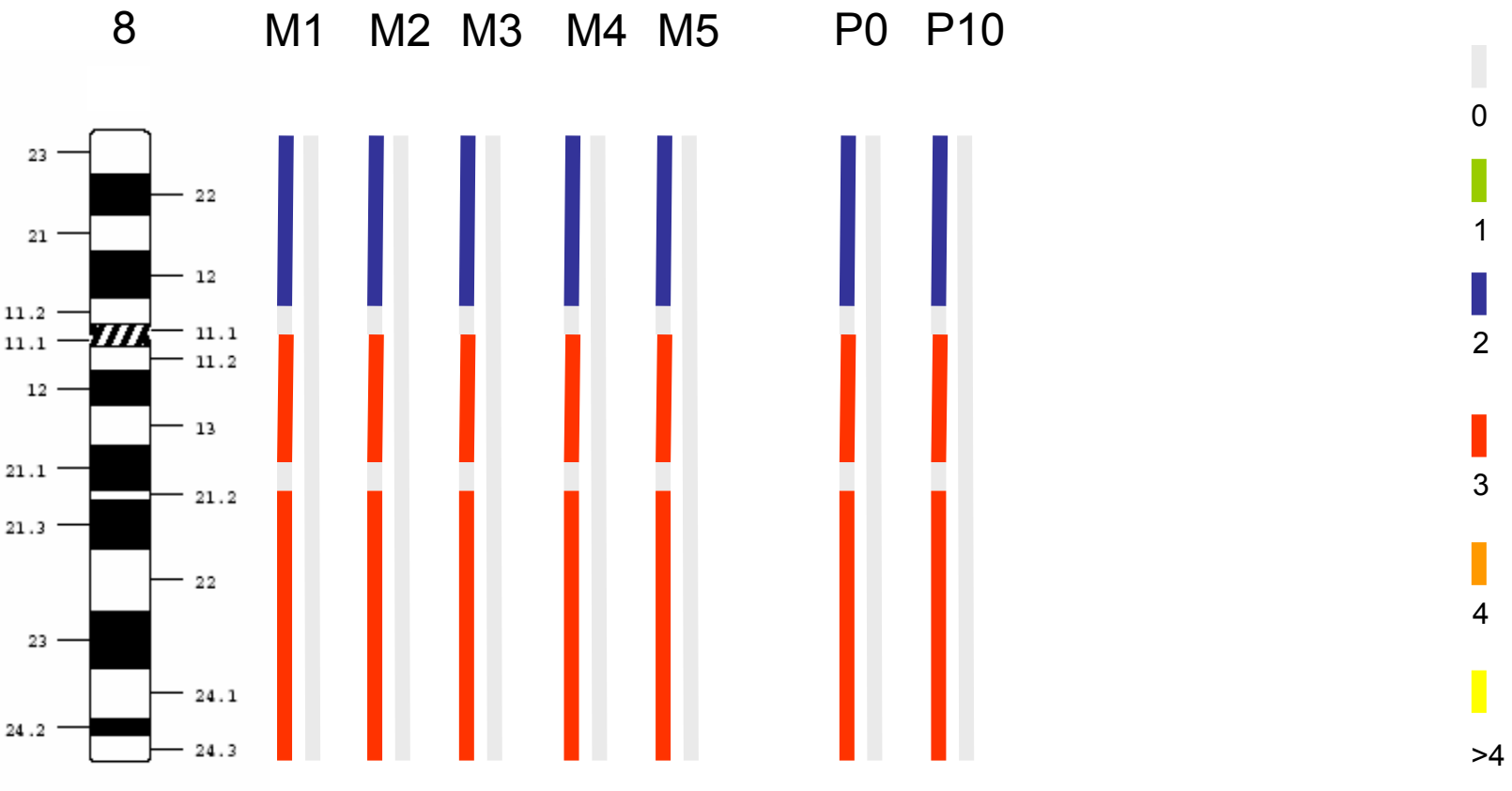

3
M1
M2
M3
M4
M5
P0
P10
26
24
22
14
12
11.1
12
13.2
21
23
25
26.2
27
29
25
23
21
13
11.1
11.2
13.1
13.3
22
24
26.1
26.3
28
0
1
2
3
4
>4

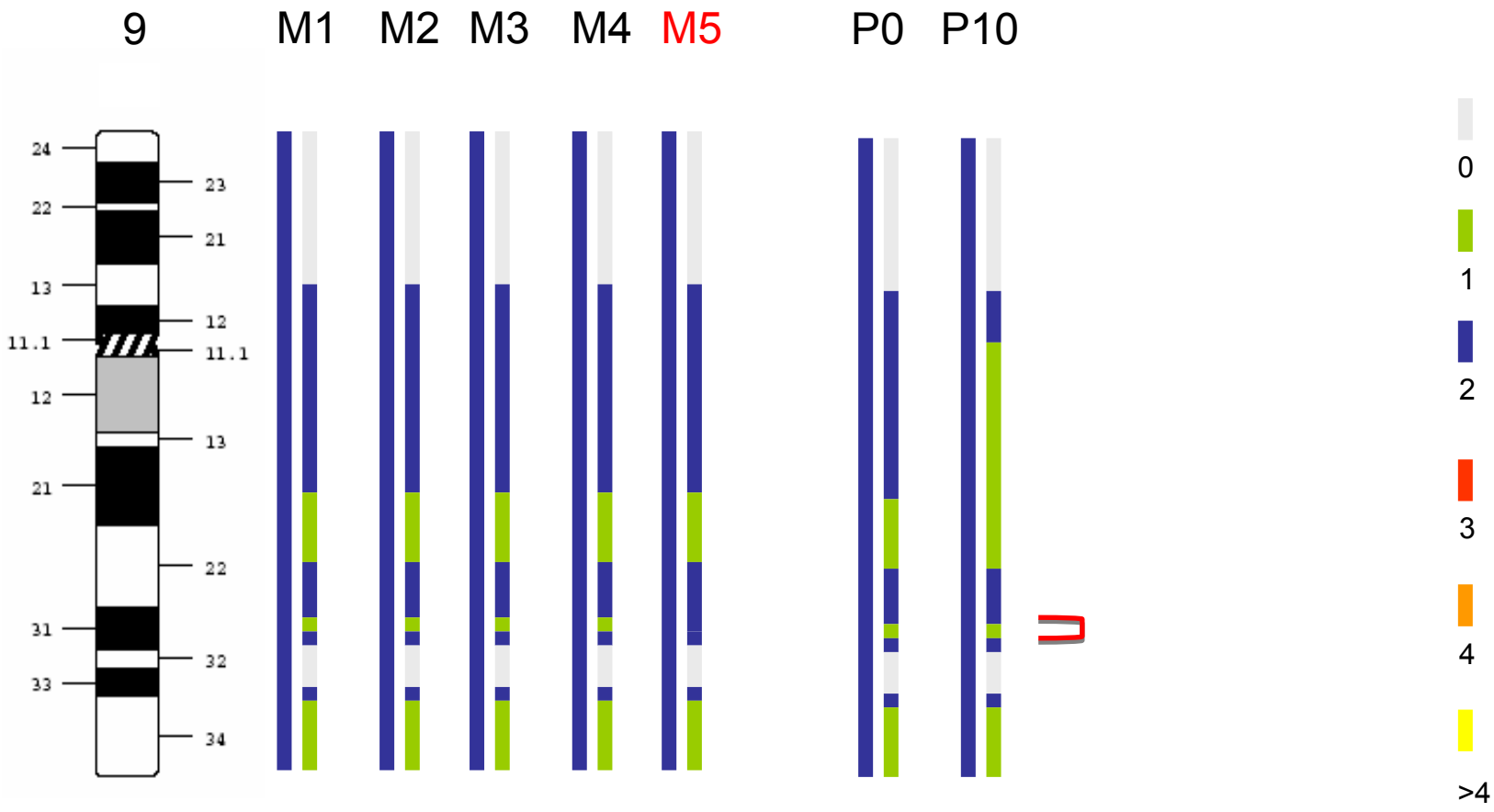

4
M1
M2
M3
M4
M5
P0
P10
16
15.3
15.2
15.1
14
13
12
11
11
12
13
21
22
23
24
25
26
27
28
31.1
31.2
31.3
32
33
34
35
0
1
2
3
4
>4

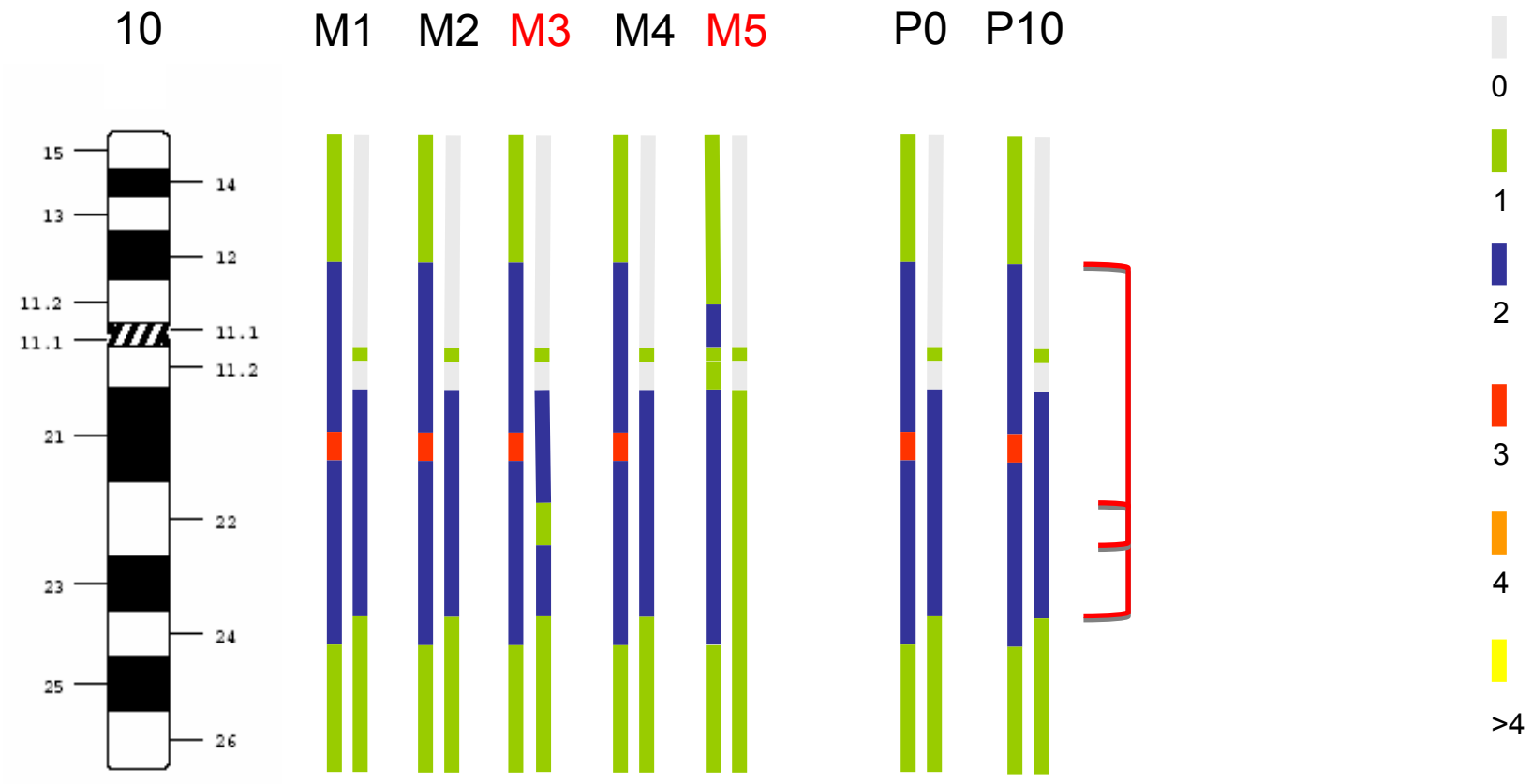

5
M1
M2
M3
M4
M5
P0
P10
15.3
15.1
13
11
11.2
13
15
22
31
33
35
15.2
14
12
11.1
12
14
21
23
32
34
0
1
2
3
4
>4

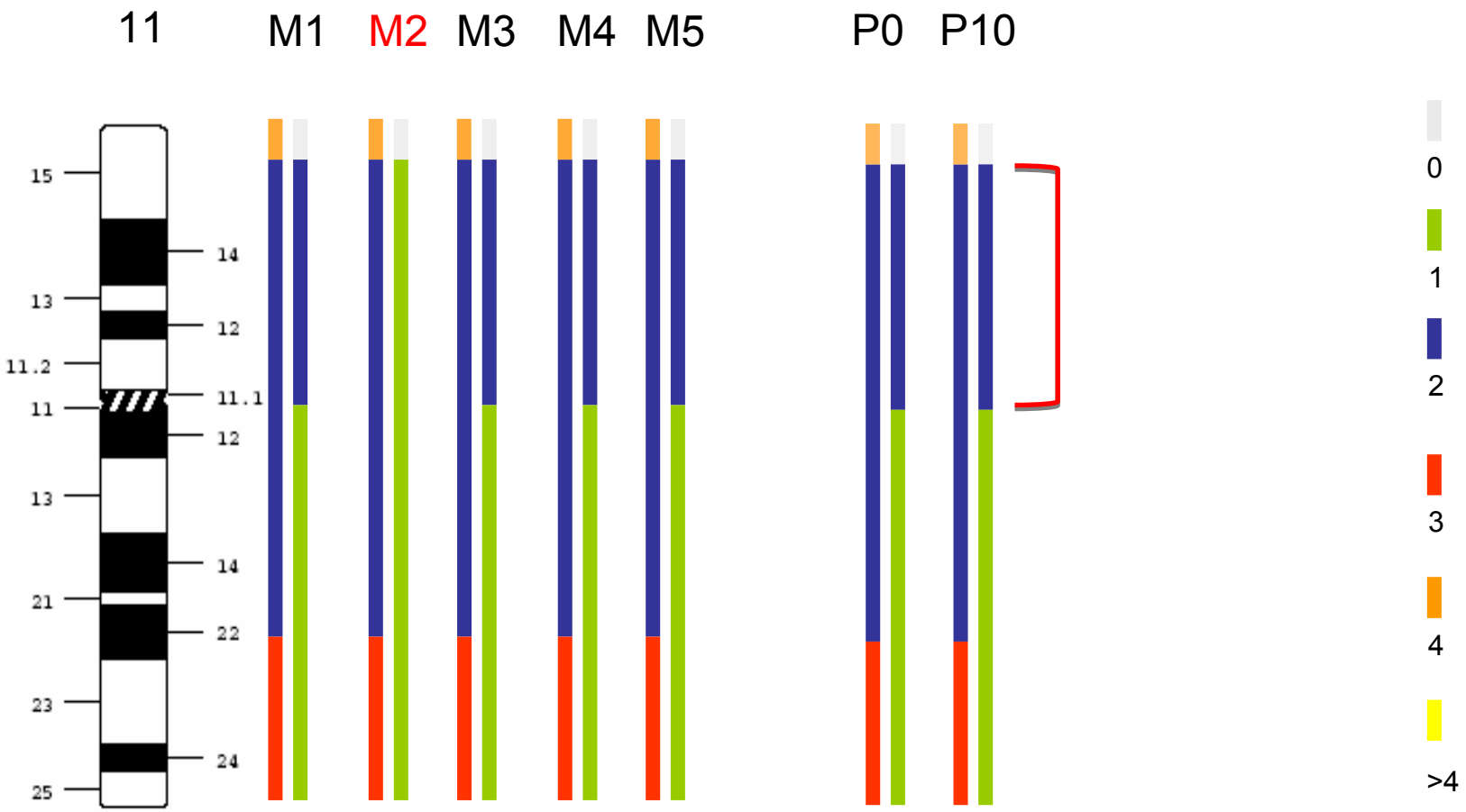

6
M1
M2
M3
M4
M5
P0
P10
25
23
21.3
21.1
11.2
11.1
13
15
21
23
25
27
24
22
21.2
12
11.1
12
14
16
22
24
26
0
1
2
3
4
>4

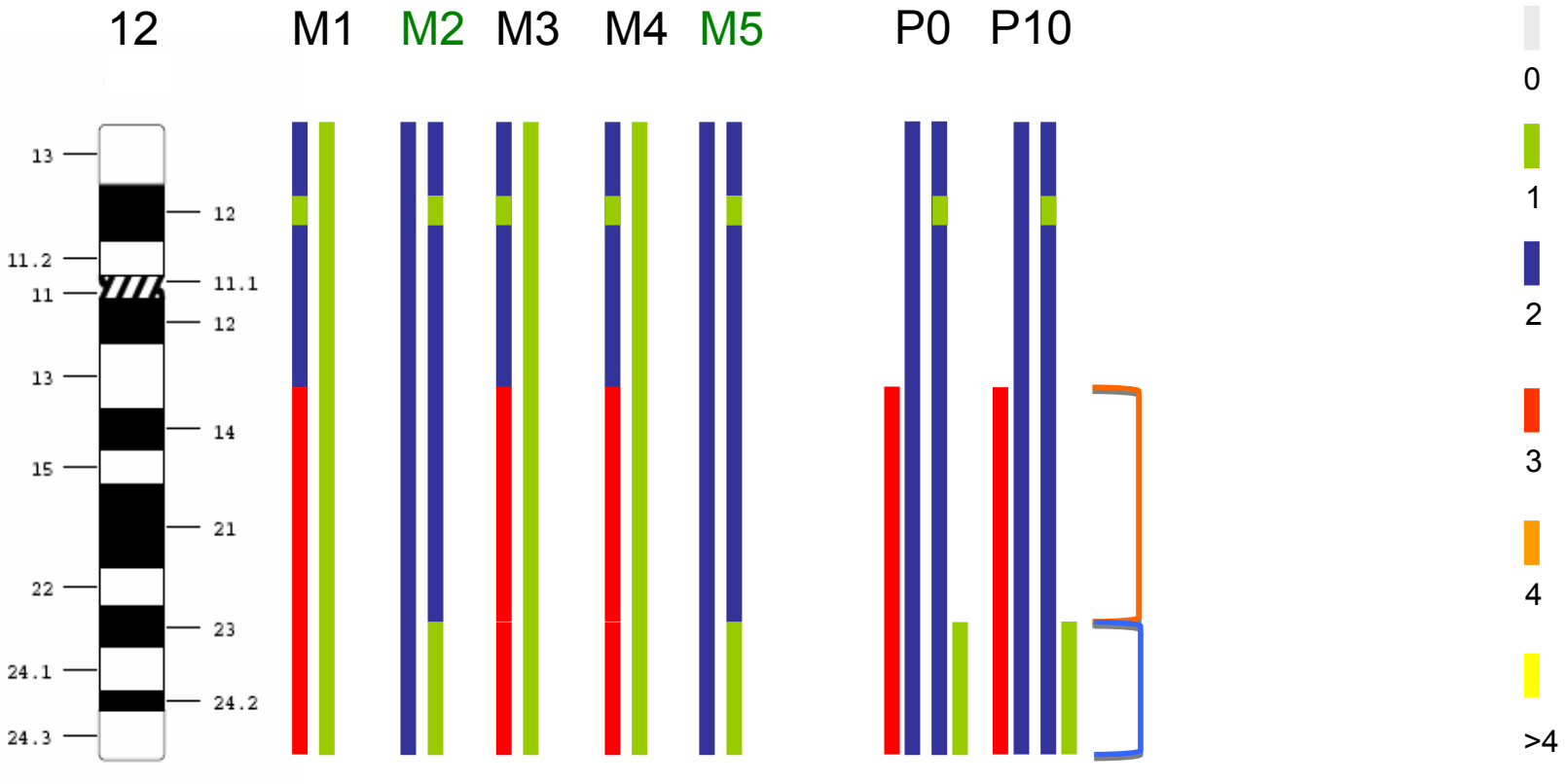

7
M1
M2
M3
M4
M5
P0
P10
22
15
13
11.2
11.1
21
14
12
11.1
11.2
21
22
31
32
33
34
35
36
0
1
2
3
4
>4

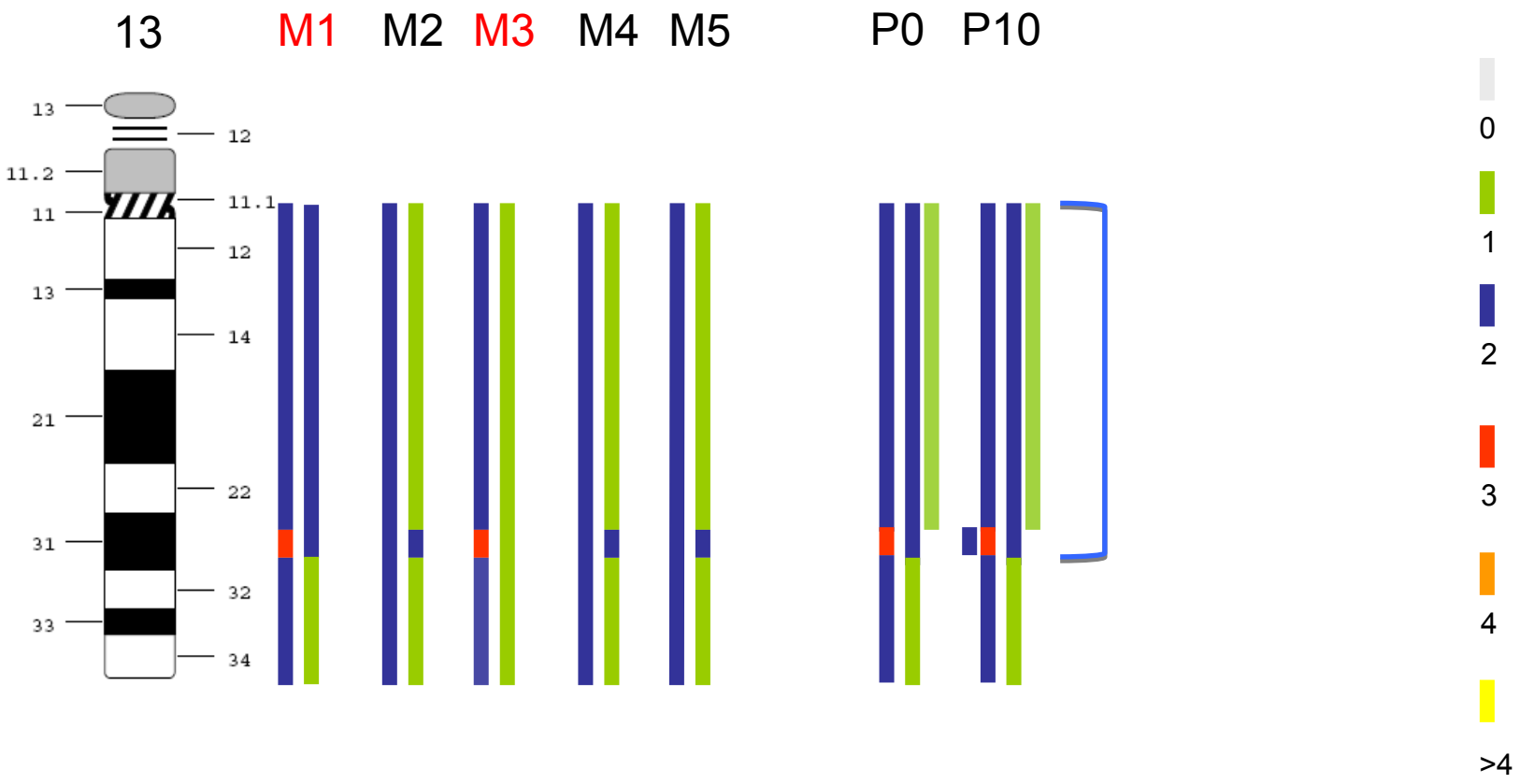

8
M1
M2
M3
M4
M5
P0
P10
23
21
11.2
11.1
12
21.1
21.3
23
24.2
22
12
11.1
11.2
13
21.2
22
24.1
24.3
0
1
2
3
4
>4

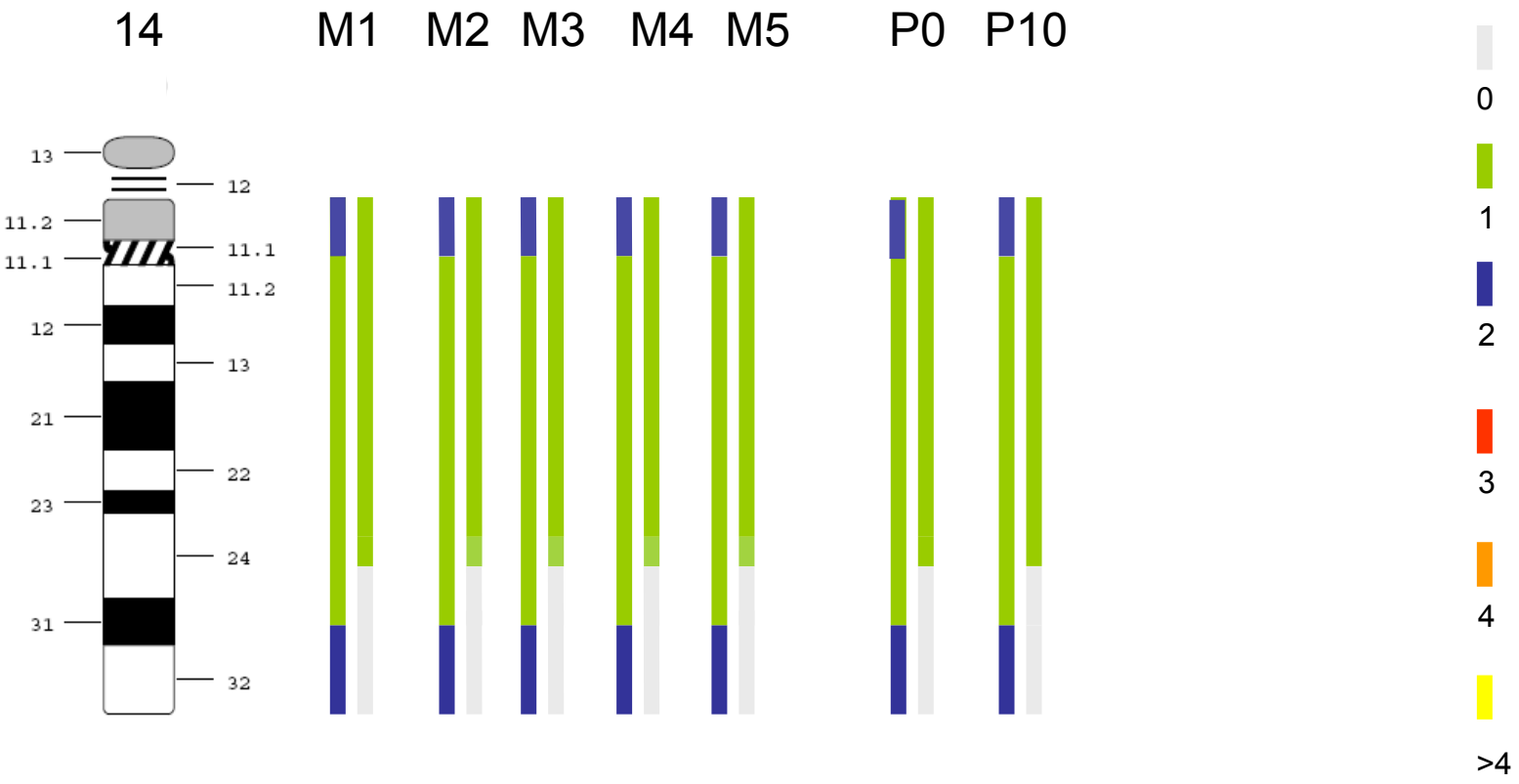

9
M1
M2
M3
M4
M5
P0
P10
24
22
13
11.1
12
21
31
33
23
21
12
11.1
13
22
32
34
0
1
2
3
4
>4

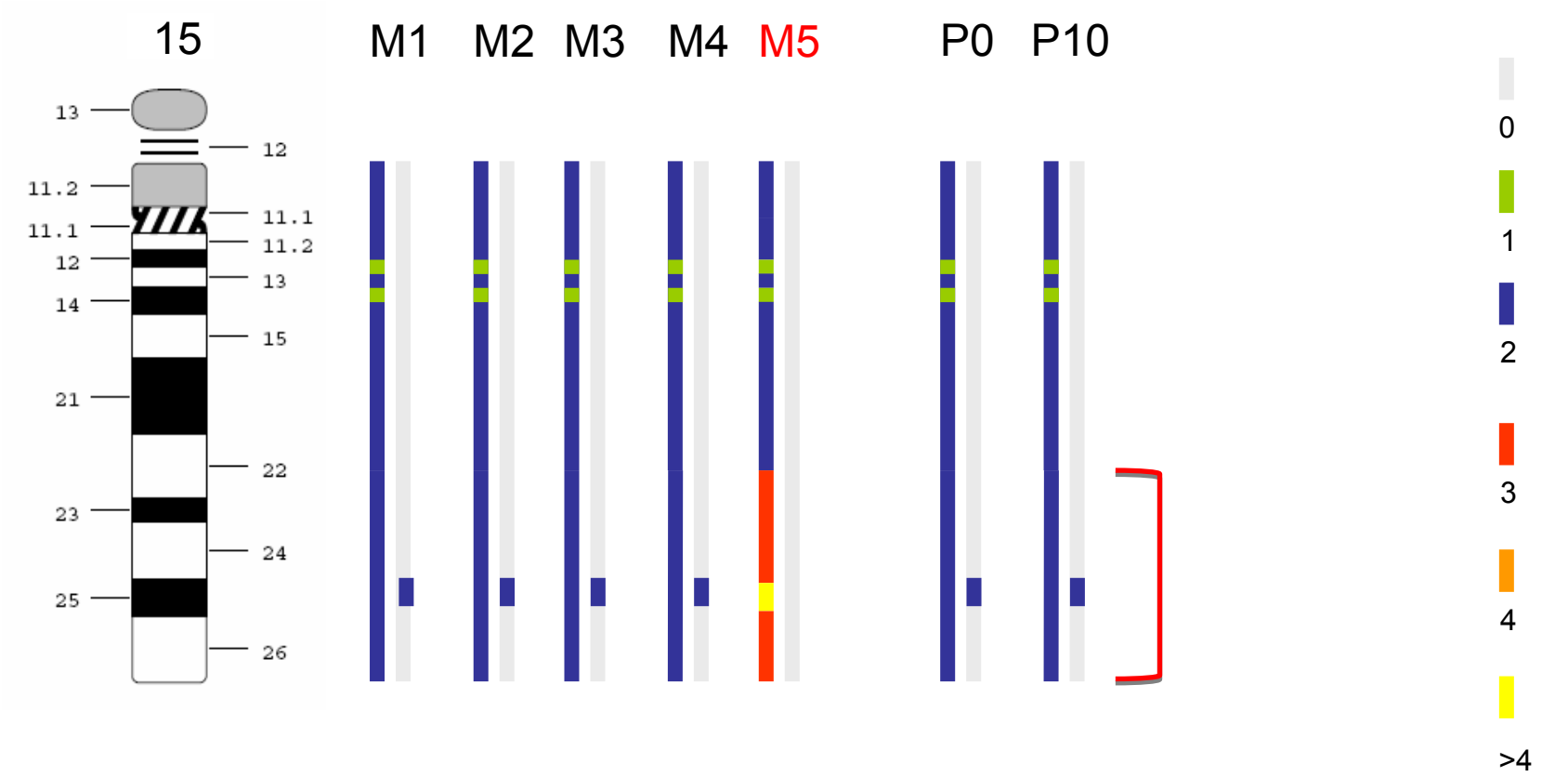

10
M1
M2
M3
M4
M5
P0
P10
15
13
11.2
11.1
14
12
11.1
11.2
21
22
23
24
25
26
0
1
2
3
4
>4

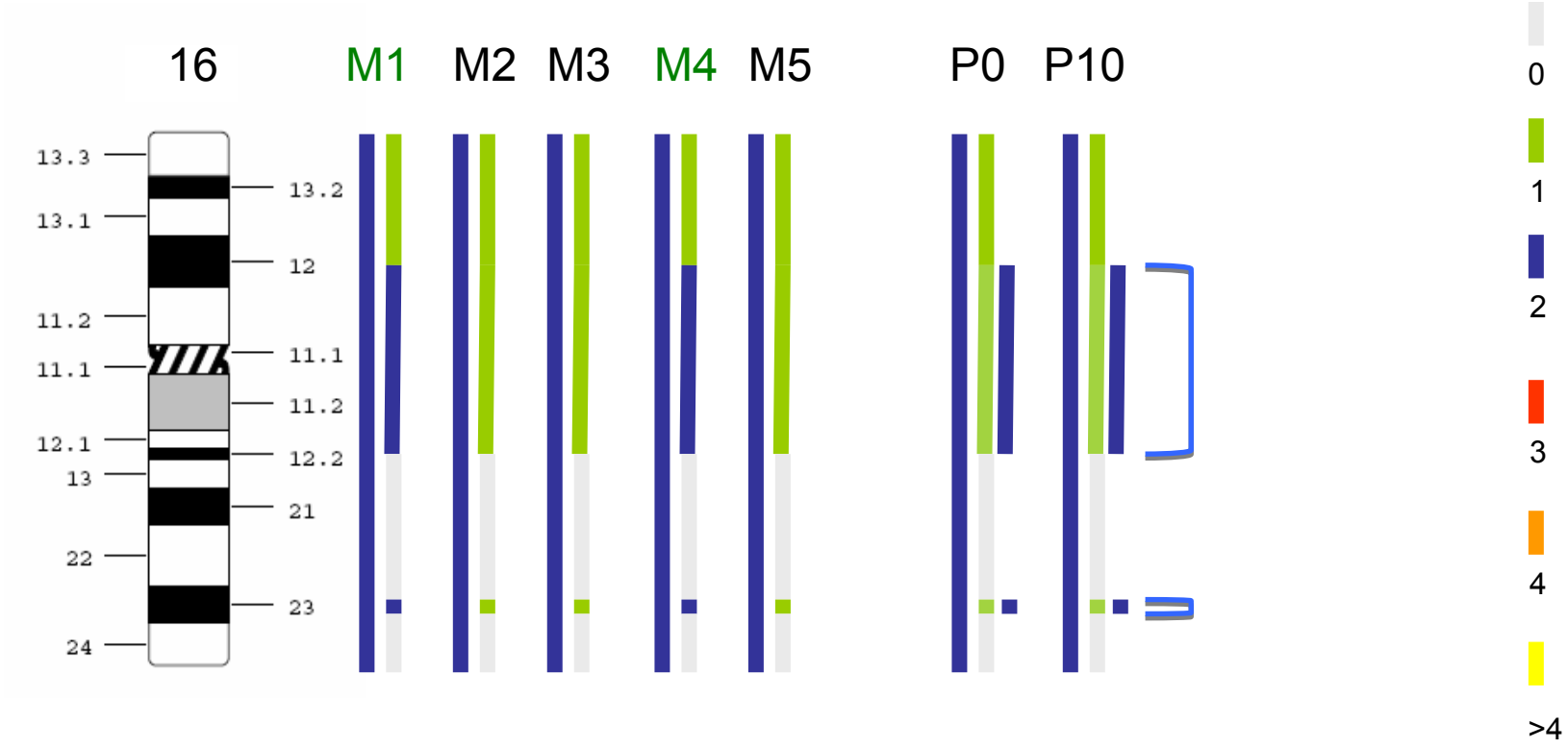

11
M1
M2
M3
M4
M5
P0
P10
15
13
11.2
11
13
21
23
25
14
12
11.1
12
14
22
24
0
1
2
3
4
>4

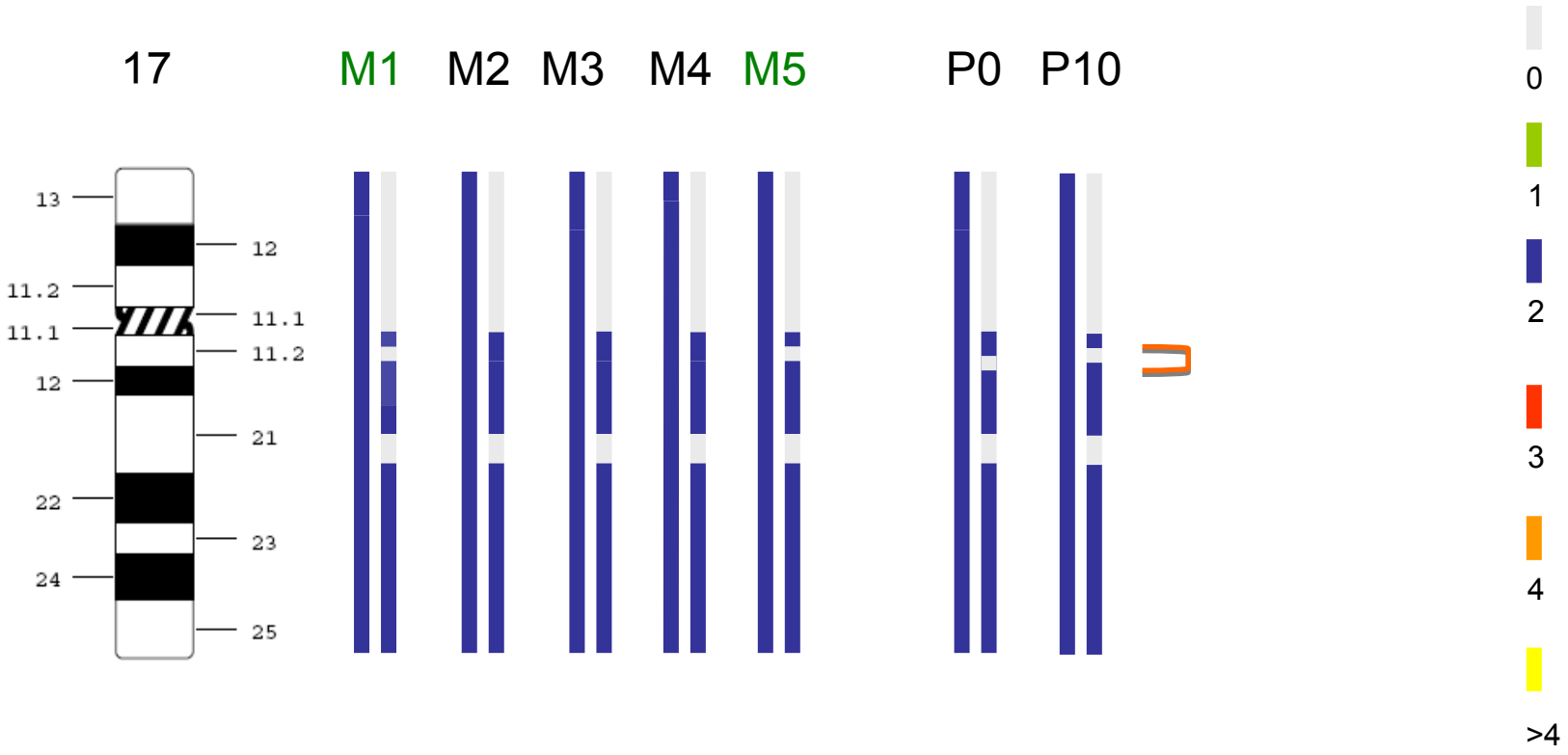

12
M1
M2
M3
M4
M5
P0
P10
13
11.2
11
13
15
22
24.1
24.3
12
11.1
12
14
21
23
24.2
0
1
2
3
4
>4

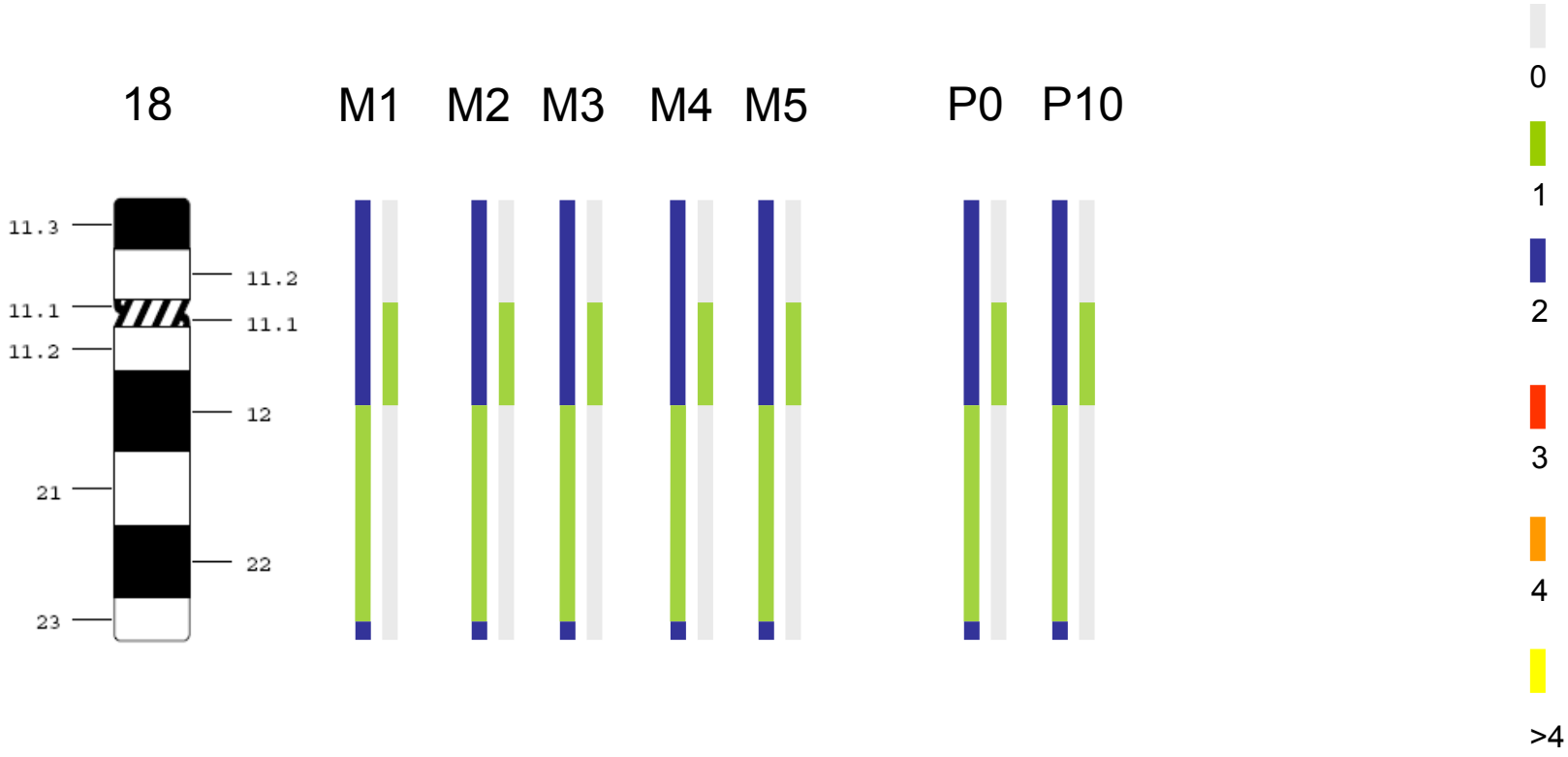

13
M1
M2
M3
M4
M5
P0
P10
13
11.2
11
13
21
31
33
12
11.1
12
14
22
32
34
0
1
2
3
4
>4

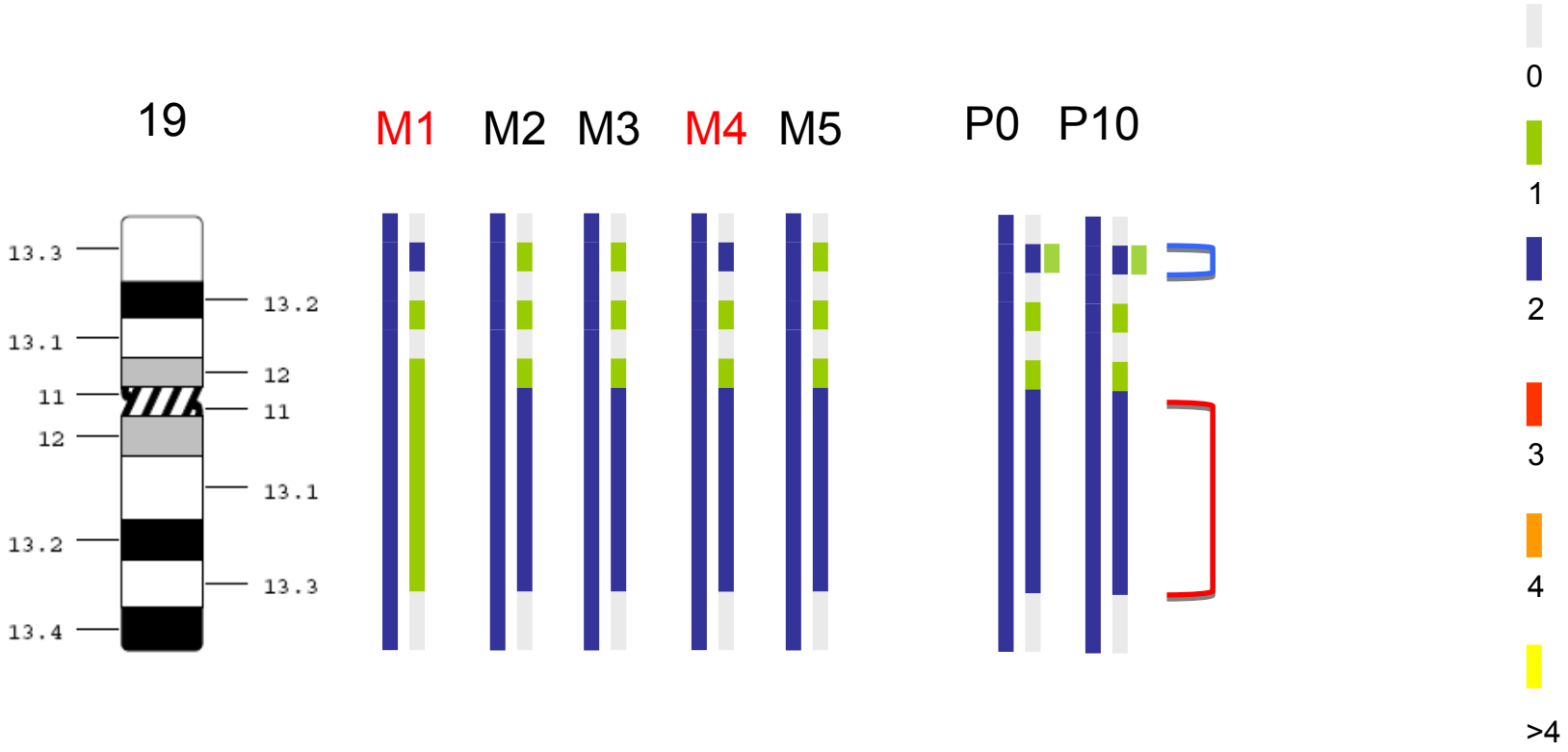

14
M1
M2
M3
M4
M5
P0
P10
13
12
11.2
11.1
11.1
11.2
12
13
21
22
23
24
31
32
0
1
2
3
4
>4

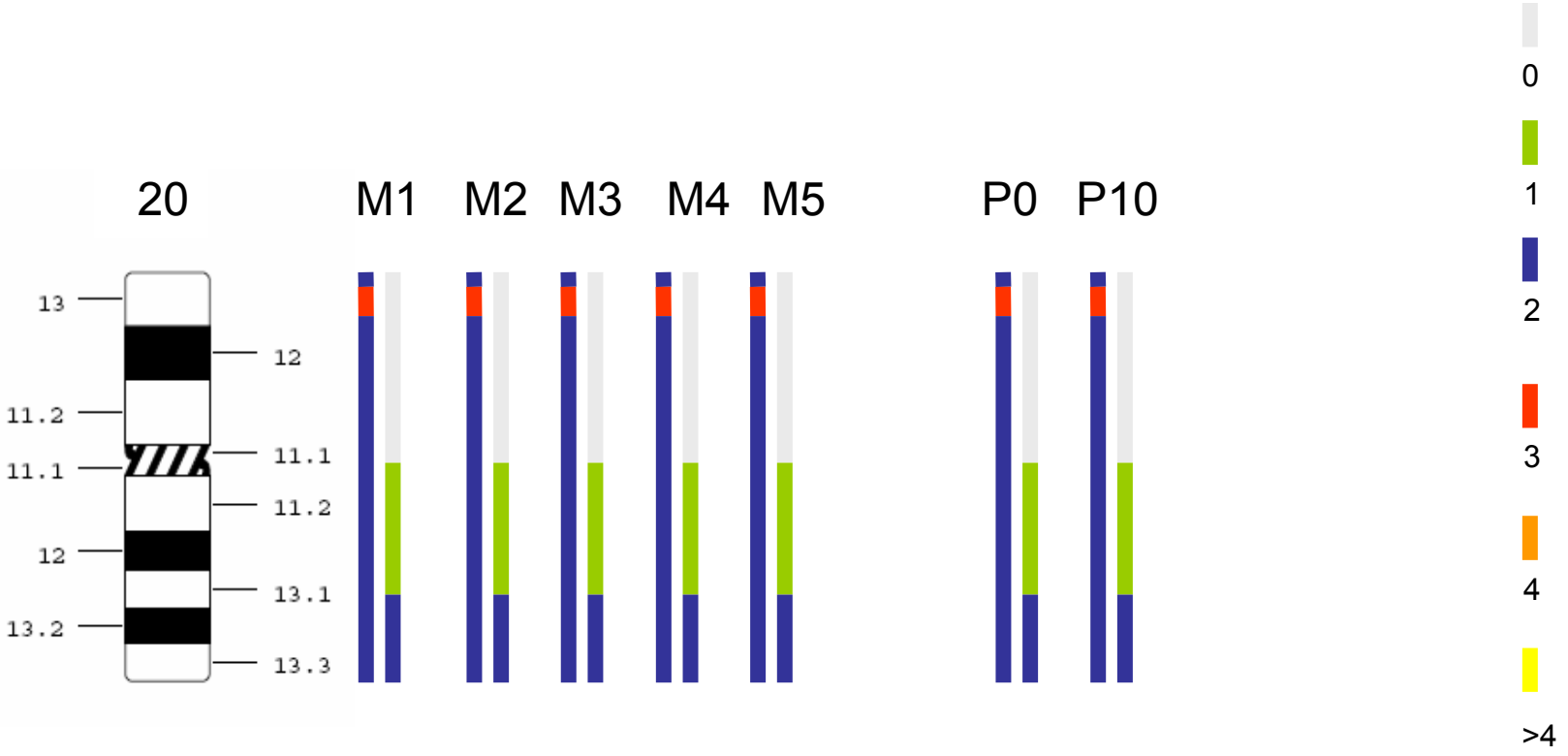

15
M1
M2
M3
M4
M5
P0
P10
13
12
11.2
11.1
11.1
11.2
12
13
14
15
21
22
23
24
25
26
0
1
2
3
4
>4

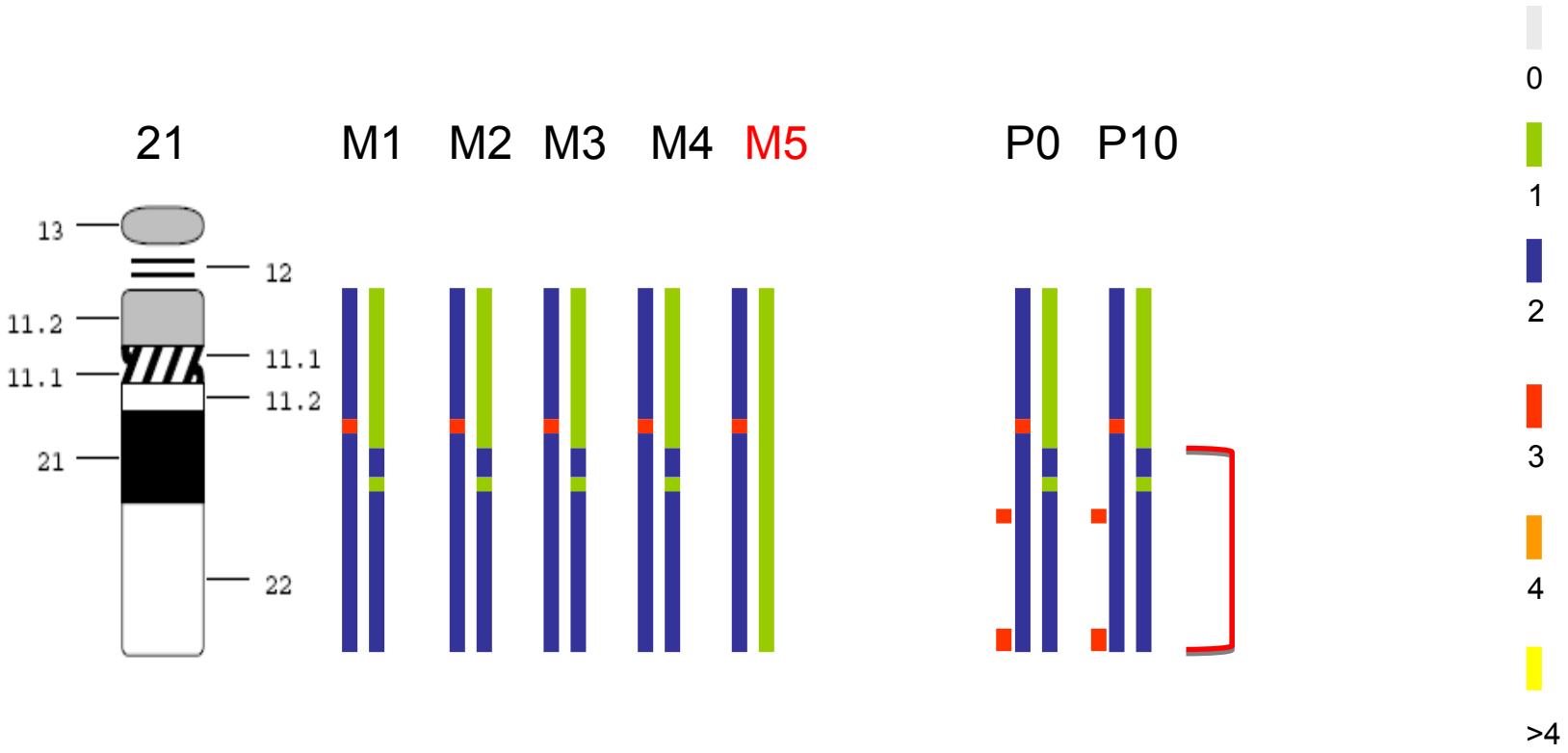

16
M1
M2
M3
M4
M5
P0
P10
13.3
13.1
11.2
11.1
12.1
13
22
24
13.2
12
11.1
11.2
12.2
21
23
0
1
2
3
4
>4

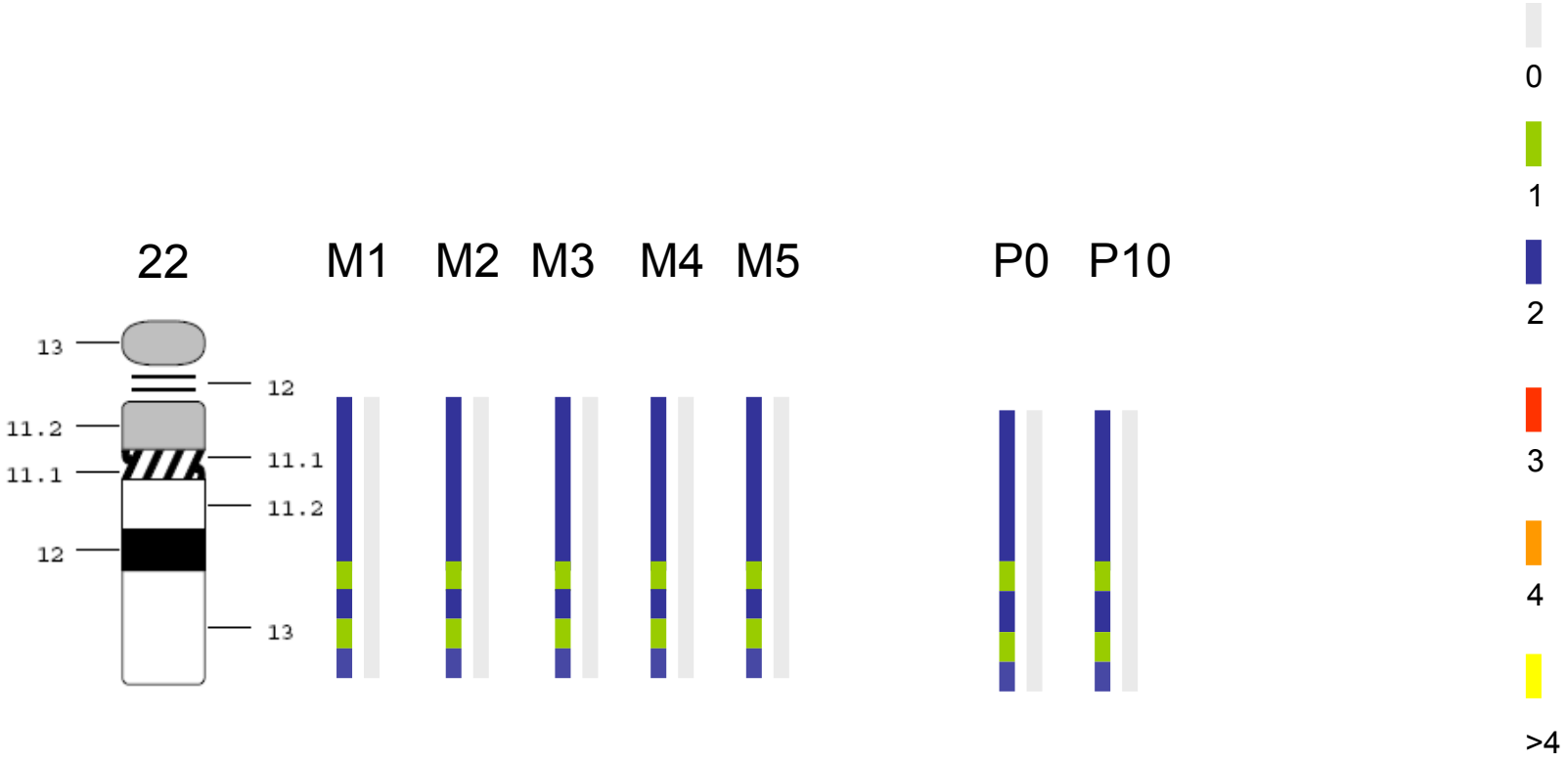

17
M1
M2
M3
M4
M5
P0
P10
13
12
11.2
11.1
11.1
11.2
12
21
22
23
24
25
0
1
2
3
4
>4

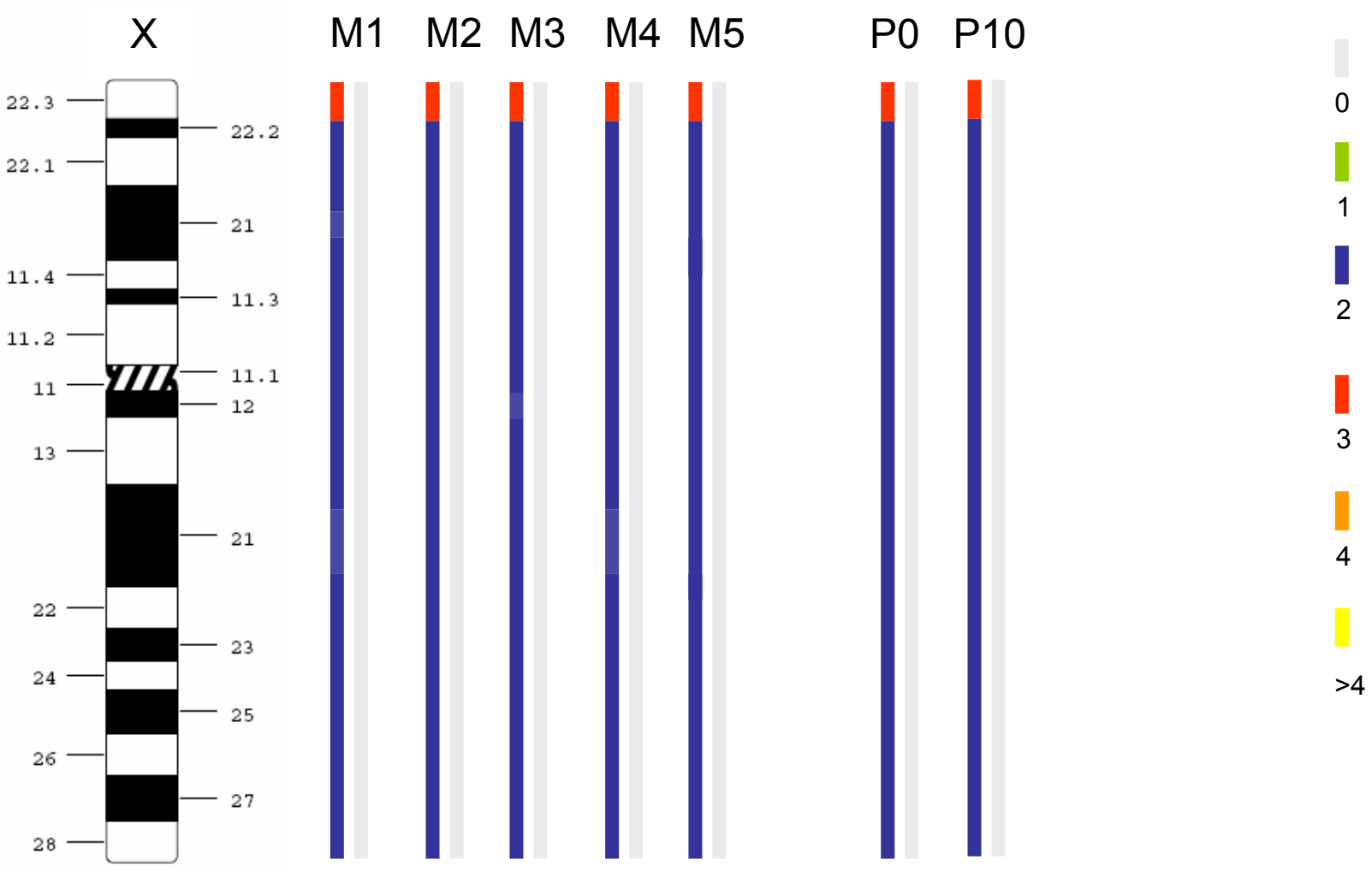

18
M1
M2
M3
M4
M5
P0
P10
11.3
11.2
11.1
11.1
11.2
12
21
22
23
0
1
2
3
4
>4

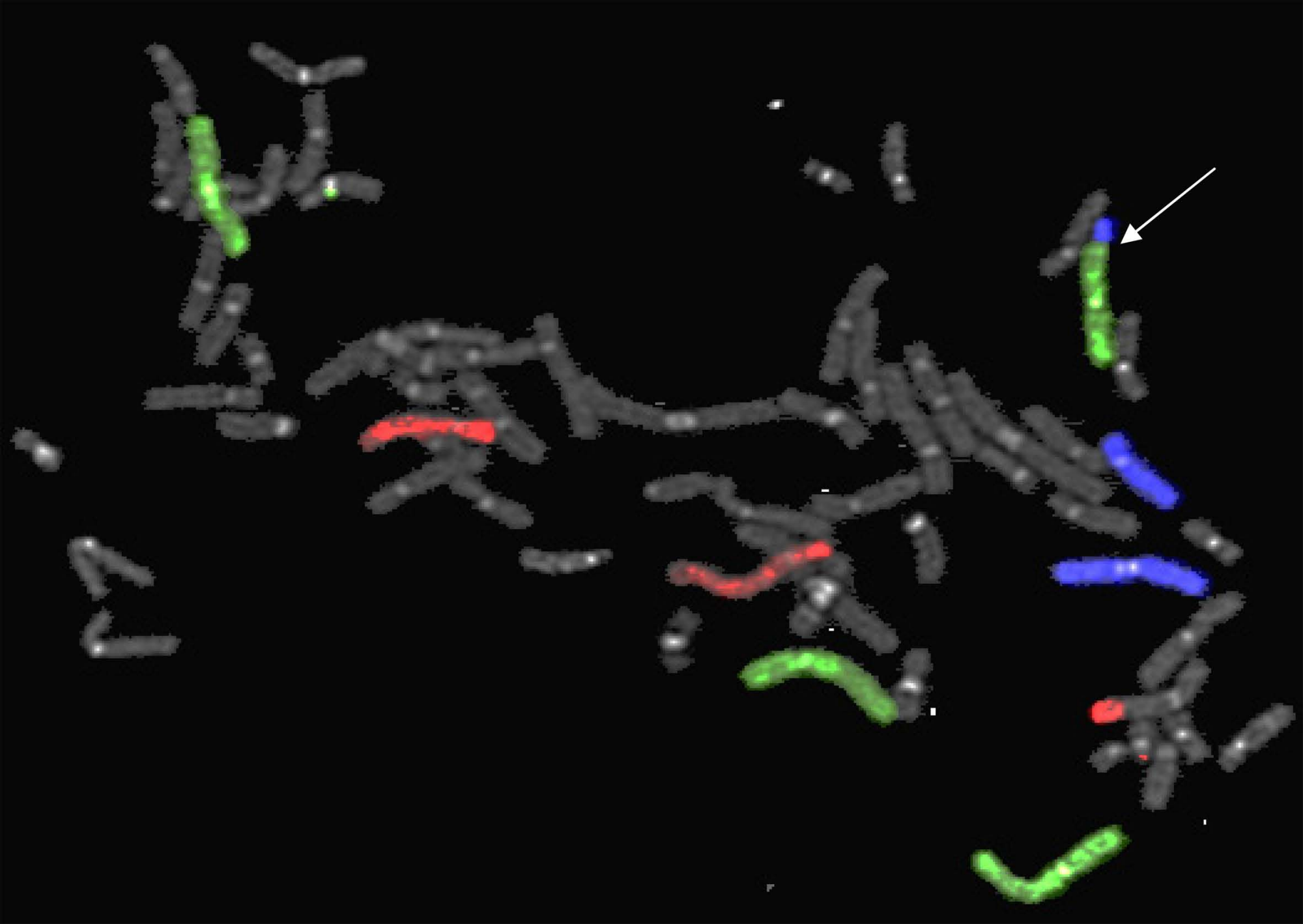

19
M1
M2
M3
M4
M5
P0
P10
13.3
13.1
11
12
13.2
13.4
13.2
12
11
13.1
13.3
0
1
2
3
4
>4


20
M1
M2
M3
M4
M5
P0
P10
13
12
11.2
11.1
11.1
11.2
12
13.1
13.2
13.3
0
1
2
3
4
>4


21
M1
M2
M3
M4
M5
P0
P10
13
12
11.2
11.1
11.1
11.2
21
22
0
1
2
3
4
>4


22
M1
M2
M3
M4
M5
P0
P10
13
12
11.2
11.1
12
11.1
11.2
13
0
1
2
3
4
>4


X
M1
M2
M3
M4
M5
P0
P10
22.3
22.2
22.1
21
11.4
11.3
11.2
11.1
11
12
13
21
22
23
24
25
26
27
28
0
1
2
3
4
>4



Supplemental Table 1. Segments with variation in copy number between specific lines.[a]

	P0	P10	M1	M2	M3	M4	M5
P0	-	9q11-q21	1q32-qter 12pter-q23 19q11-q13	1p36-q24 1q32-q41 1q42-qter 5q32 6q14-q21 11p15-p11 13q31 17q11	1q32-qter 10q22 12 pter -q23 17q11	1q32-qter 12pter -q23 13q31 17q11	1p36-q24 1q32-q41 1q42-qter 2pter-qter 6q14-q21 9q31 10p12-q24 13q31 15q22-qter 21q11-qter
P10	9q11-q12	-	1p36-q24 1q32-q41 9q11-q21 12pter-q23 19q11-q13	5q32 9q11-q21 11p15-p11 17q11	1p36-q24 1q32-q41 9q11-q21 10q22 12pter-q23 17q11	1p36-q24 1q32-q41 9q11-q21 12pter-q23 17q11	2pter-2qter 9q11-q21 9q31 10p12-q24 15q22-qter 21q11-qter
M1	1q32-qter 12pter-q23 19q11-q13	1p36-q24 1q32-q41 9q11-q21 12pter-q23 19q11-q13	-	1p36-q24 1q32-qter 5p14 5q32 6q14-q22 11p15-q21 12pter-qter 13q11-q31 16p12-q11 16q23 17q11 19p13 19q11-q13	5p14 10q22 13q11-q31 16p12-q11 16q23 17q11 19p13 19q11-q13	13q11-q31 19q11-q13	1p36-q24 1q32- qter 2pter-qter 5p14 6q14-q22 7q32-qter 9q31 10p12-q24 12pter-qter 13q11-q31 15q22-qter 16p12-q11 16q23 19p13 19q11-q13 21q11-qter
M2	1p36-q24 1q32-q41 1q42-qter 5q32 6q14-q21 11p15-p11 13q31 17q11	5q32 9q11-q21 11p15-p11 17q11	1p36-q24 1q32-qter 5p14 5q32 6q14-q22 11p15-q21 12pter-qter 13q11-q31 16p12-q11 16q23 17q11 19p13 19q11-q13	-	1p36-q24 1q32-qter 5q32 6q14-q22 10q22 11p15-p11 12pter-qter 13q31	1p36-q24 1q32-qter 5p14 5q32 6q14-q22 11p15-p11 12pter-qter 16p12-q11 16q23 19p13	2pter-qter 5q32 7q32-qter 9q31 10p12-q24 11p15-p11 15q22-qter 17q11 21q11-qter
M3	1q32-qter 10q22 12 pter -q23 17q11	1p36-q24 1q32-q41 9q11-q21 10q22 12pter-q23 17q11	5p14 10q22 13q11-q31 16p12-q11 16q23 17q11 19p13 19q11-q13	1p36-q24 1q32-qter 5q32 6q14-q22 10q22 11p15-p11 12pter-qter 13q31	-	5p14 10q22 13q31 16p12-q11 16q23 19p13	1p36-q24 1q32-qter 2pter-qter 6q14-q22 7q32-qter 9q31 10p12-q24 12pter-qter 13q31 15q22-qter 17q11 21q11-qter
M4	1q32-qter 12pter -q23 13q31 17q11	1p36-q24 1q32-q41 9q11-q21 12pter-q23 17q11	13q11-q31 19q11-q13	1p36-q24 1q32-qter 5p14 5q32 6q14-q22 11p15-p11 12pter-qter 16p12-q11 16q23 19p13	5p14 10q22 13q31 16p12-q11 16q23 19p13	-	1p36-q24 1q32-qter 2pter-qter 5p14 6q14-q22 7q32-qter 9q31 10p12-q24 12pter-qter 15q22-qter 16p12-q11 16q23 17q11 19p13 21q11-qter

[a] Shifts between main clones and sub-clones not included

Supplemental Table 2. Number of segments exhibiting variation between specific lines.[a]

	P0	**P10**	**M1**	**M2**	**M3**	**M4**	**M5**
P0	-	1	3	8	4	4	10
P10	1	-	5	4	6	5	6
M1	3	5	-	13	8	2	16
M2	8	4	13	-	8	10	9
M3	4	6	8	8	-	6	12
M4	4	5	2	10	6	-	15

[a] Shifts between main clones and sub-clones not included