CHAPTER 7. GENOME-WIDE MISMATCH REPAIR SCREEN

7.1 Introduction

The data of the pilot screen described in Chapter 6 indicated that cells with mutations in mismatch repair (MMR) genes could be identified in a *Blm*-deficient ES cell mutation library generated by irradiation. It also showed that array CGH with 200 kb resolution was able to scan the mouse genome for deletions induced by irradiation. Thus, in this chaper a new mutation library with higher genome coverage was generated and screened (Figure 7-1). In this screen, we expected to identify mutants of known MMR genes and discover new ones. Prior to generating the mutation libraries, the cells were tagged with a recombinant MMuLV retrovirus to avoid parallel analysis of daughter clones. This is particularly important because array analysis is very expensive and it is much better to analyse more clones than the same clone multiple times. The diverse genomic locations of these molecular tags facilitate the identification of relationships between clones in the same pool. Seven mutation pools were generated by irradiating the MMuLV tagged *Blm*-deficient cells (NGG5-3) at 10 Gray. Screening these individual pools has at least two advantages: Firstly, it is easy to distinguish independent mutants among pools. If a gene is very important to the MMR system and its mutants are able to be isolated, mutants of this gene can be expected in more than one pool due to the high genome coverage of irradiation-induced mutations. Mutation pools were generated and screened separately, thus clones in one pool are independent from clones in another pool. Secondly, individual mutation pools prevent one mutant from dominating the whole mutation library. The screen is targeted for homozygous deletions generated through loss of heterozygosity, which may occur at anytime in cell culture. Therefore, mutants for which LOH occurred earlier can be many times more abundant than mutants for which LOH occurred later. A single mutant with an early LOH event generating a homozygous mutation can replicate to be approximately 1000 (2^{10})–10,000 (2^{14}) times more than any other mutants. By using several mutation pools, this phenomenon, though it occurs, will not affect the process of isolating mutants in other pools.

Based on the previous data of irradiation survival frequency, the number of cells to be irradiated can be calculated to achieve a required number of independent mutants. This must be less than the total number of the tags in each pool. This ensures that each surviving clone has a unique tag and can be distinguished. It is known from Chapter 5 that two-thirds of surviving cells contain deletions and each surviving cell has ~10 Mb deleted regions following a 10 Gray irradiation. The haploid mouse genome is 3,400 Mb in length

thus one thousand surviving clones in a pool ensure that deletions can cover the mouse genome twice. After irradiation, the mutant pools were expanded for 15 cell doublings which leaves enough time for non-sister homologous chromatid exchange to generate LOH events and also to allow the decay of messenger RNA and proteins of mutated MMR genes. The mutation library was then selected in 6TG media to isolate candidate MMR-deficient clones.

Array CGH analysis is able to detect genomic heterozygous and homozygous deletions generated by irradiation. Homozygous deletions at known MMR genes and novel genes are anticipated in the mutants. Observation of homozygous deletions at known MMR loci, for instance *Msh2* and *Msh6*, would confirm the reliability of the screen system. At the same time, new mutant genes should be identified. However, heterozygous deletions, recognized by array CGH, are also worth further consideration as they may lead to loss-of-function mutations. Small mutations, occurring in the other non-deleted homologous DNA sequence of the so-called heterozygous deletion, can also cause loss-of-function mutations, such as frame shift mutants.

In the mutation library induced by irradiation, MMR mutants may have mutations, such as small mutations or translocations, which cannot be detected by the 200 kb array CGH, but these can result in abnormality at the transcriptional level. Thus, mutant clones were also analysed using the Illumina® Mouse-6 Expression BeadChip. Small mutations can cause loss-of-function mutations through the mechanism of nonsense-mediated decay (NMD). NMD is an RNA quality surveillance pathway, by which mRNAs harbouring premature termination (nonsense) codons (PTC) can be selectively degraded (Chang *et al.* 2007). The mRNA molecules with PTCs can generate deleterious functions if they are translated to truncated proteins. It is necessary to remove these mRNAs to avoid deleterious functions. Irradiation can introduce premature termination codons in mRNA through reading frame shifts. The absence of its mRNA led by NMD can be detected by expression array. Moreover, reading frame shifts caused by translocations can also result in PTCs and the degradation of its mRNA through the NMD pathway. Another application of expression array analysis is the identification of genes which are regulated directly or indirectly by defects in MMR.

To summarise, large homozygous and heterozygous deletions can be identified by array CGH; the silenced genes caused by smaller mutations and translocations can also be detected by the expression array. By combining two array analysis approaches, novel MMR genes and components in the MMR-associated pathway can be identified.

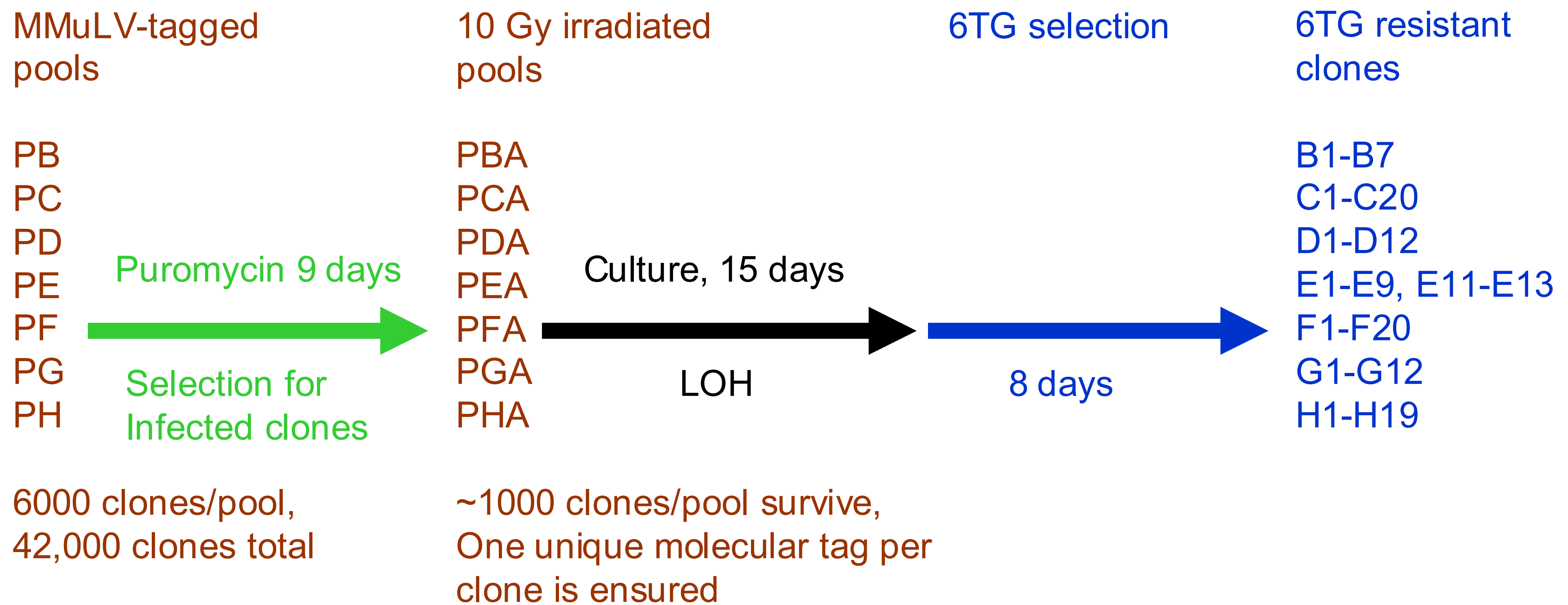

Figure 7-1 **The procedure of a genome-wide MMR screen on the MMuLV-infected, irradiated ES cell library**

A recombinant retroviral vector (MMuLV) containing the *PuroΔtk* selection marker was used to infect *Blm*-deficient ES cells in seven pools (PB–PH). Cells were selected in puromycin for nine days. The ratio of virus to cells was controlled to ensure that each cell was only infected once. Each pool consisted of 6,000 independently infected cells. When these cells grew to form colonies, they were trypsinized and irradiated separately, which resulted in ~1000 clones in each pool. These were expanded for 15 days in non-selective media and then selected in 6TG media. The resulting MMR-deficient candidates are termed B–H followed by clone numbers.

7.2 Results

7.2.1 Use of MMuLV to provide molecular tags in mutation pools

The *Blm*-deficient ES cells (NGG5-3) were infected in seven pools (PB–PH) by a recombinant MMuLV retrovirus containing the *PuroΔtk* cassette. The ratio of recombinant virus to ES cells was controlled so that each cell was infected only once. The recombinant MMuLV was generated by a virus producing cell line B4-5 (Wang and Bradley 2007). This cell line was cultured in non-selective ES cell media for one day then the media was fitered to remove any cells. When the cells were 70–80% confluent, they are trypsinized and plated with an appropriate volume of media containing the virus. Infected cells were selected in puromycin and the resisting colonies composed of mutation pools. The viral titer was determined by plating three million cells with 60μL media containing the virus, followed by puromycin selection. The resulting colonies were stained by methylene blue and counted. Under this condition, 607 infected colonies were identified, which gives a titre of 10,000 viral particles per mL. The 600μL media containing the virus was used to infect three million *Blm*-deficient cells per pool and approximately 6000 colonies per pool were achieved. Each pool were irradiated at 10 Gray to generate mutation pools according to the data in Chapter 5, survival frequency of the *Blm*-deficient cells irradiated at 10 Gray was ~0.1%. Therefore, I expect that mutation pool has a complexity of 800 clones.

Given that there were 6000 independent retroviral insertions per pool before cells were expanded for irradiation. Thus, in the 800 colonies of each pool after the irradiation, each should carry a unique molecular tag provided by the recombinant MMuLV. Thus, clonal relationships within a pool can be determined by Southern blot analysis using a *PuroΔtk* probe and an enzyme that generates a junction fragment between the provirus and the genomic DNA. If the fragments identified by Southern blot analysis are the same between two $6TG^R$ clones, it is highly likely that they are derived from the same clone. Relationships can be verified later by the pattern of genomic alterations.

7.2.2 Gamma-irradiation mutation library

A dose of 10 Gray was used to irradiate 4×10^6 cells per pool. Cells were plated in one 90mm plate per pool. This resulted in around one thousand independent mutant clones per pool, which is the complexity of the pool (Table 7-1).

Each mutant clone formed a colony after 10 days' culture in non-selective media. Colonies of each pool were trypsinized, passaged and cultured for another five days in non-selective

media. Then mutation pools were ready for 6TG selection as 15 cell doublings should generate at least one LOH event of each mutation. Ten million cells of each mutation pool were selected in 6TG at the density of two million cells per 90 mm plate.

Table 7-1 Complexity of each mutation pool

	Pools						
	PB	**PC**	**PD**	**PE**	**PF**	**PG**	**PH**
Colony number (10^5 cells/well)	21	19	20	19	25	33	26
Survival frequency	0.021%	0.019%	0.020%	0.019%	0.025%	0.033%	0.026%
Pool complexity	840	760	800	760	1000	1320	1040

Cell pools PB–PH were irradiated at 10 Gray. After irradiation, 100,000 cells were plated in a well to calculate irradiation survival frequency for each mutation pool. According to these data, about 1000 (760–1320) independent clones were obtained in each mutation pool.

7.2.3 Assessment of screen background

In order to control for the presence of existing MMR mutants in the *Blm*-deficient parental cells (NGG5-3) used to construct the irradiation library, the cells were plated in 6TG under the conditions used to screen the library pools (Figure 7-2). This cell density was selected for two reasons: 1. This number of cells will not grow to confluency during the screen process; 2. This cell density will select against *Hprt*-deficient mutants. *Hprt*-deficient ES cells are resistant to 6TG, however, at this high cell density, surrounding *Hprt*-proficient cells can pass 2'-deoxy-6-thioguanosine-triphosphate, a product with a phosphoribosyl group transferred by hypoxanthine guanine phosphoribosyl transferase, into the *Hprt*-deficient cell, thus killing the *Hprt*-deficient cells. Therefore, *Hprt*-deficient mutant clones can be removed from the $6TG^R$ mutant clones. In *Hprt*-proficient cells, 6TG can integrate into DNA strand as a nucleotide analogue and be methylated by an intracellular methyl group donor, S-adenosylmethionine (SAM) and forms S^6-methylthioguanine [^{SMe}G] (Karran *et al.* 2003; Swann *et al.* 1996). The resulting mismatch is then recognized by MutSα (MSH2 and MSH6) and the cell is led to cell cycle arrest and apoptotic cell death. When an MMR component, such as MSH2 protein, is deficient, cells survive in 6TG.

To measure the background of this MMR screen, four million parental *Blm*-deficient cells were selected in 2 μM 6TG at the density of 2×10^6 cells per plate for 8 days. These cells were cultured for another five days in non-selective media (M15) to allow potentially surviving clones to form bigger colonies. These two plates were stained by methylene blue to visualize colonies. A single colony-like blue spot was observed in one plate and none was observed in the other plate. According to this result, a maximum of 2–3 pre-existing MMR mutant might be isolated in each mutation pool. If a similar number of clones were recovered from one pool, these clones would possibly be false-positives. However, since the mutations in each pool were designed to cover the mouse genome three times and each heterozygous mutation can lose its heterozygosity in 12 cell doublings ($-Log_2(4\times10^{-4})$), in theory, each original mutant can have less than 8 (2^3) homozygous mutant clones after 15 cell doublings. Therefore the real positive clones in each pool should be numerous, thus this level of background should not interfere with the results.

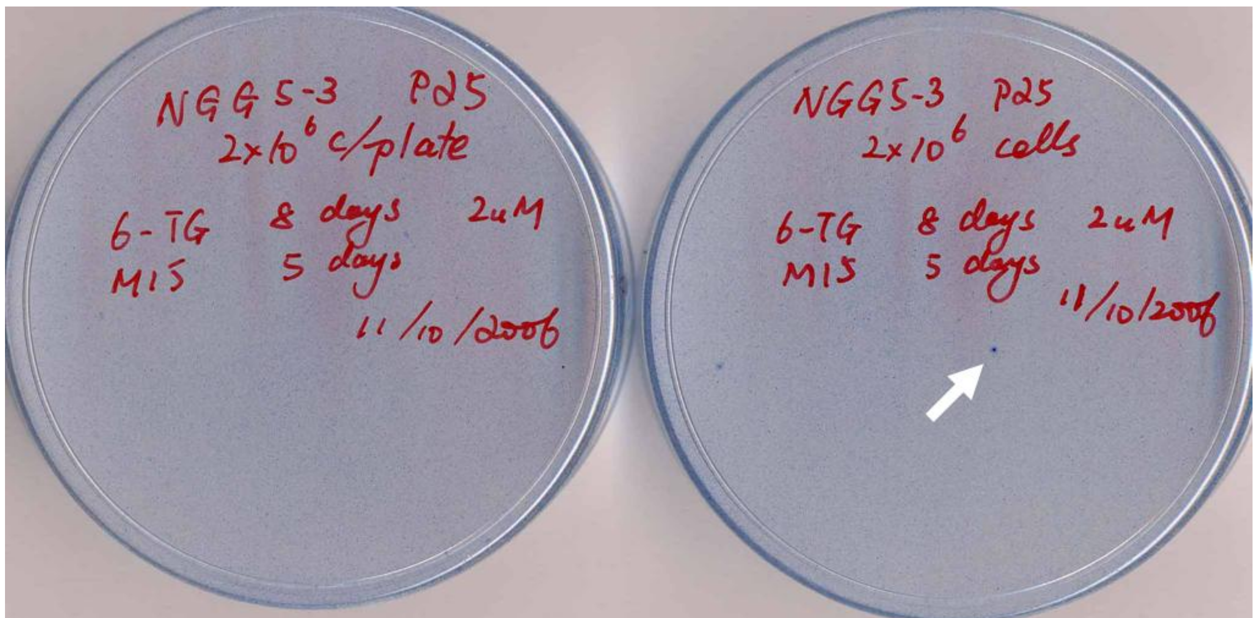

Figure 7-2 Clean background of the 6TG screen on *Blm*$^{-/-}$ cells

Two million NGG5-3 *Blm*-deficient cells were plated in each 90 mm plate. After 24 hours, cells were selected in 2 μM 6TG for eight days and recovered in M15 normal media for another five days. Then plates were stained by methylene blue. Only one colony-like blue spot is in both plates (indicated by a white arrow).

7.2.4 6TG-resistant clones

In total, 103 $6TG^R$ clones were identified from the seven pools. These were isolated in 96-well plates and named B–H followed by a clone number. These clones were distributed between 7 mutant pools with 7–20 clones per pool (Table 7-2). The 6TG resistance of each clone was re-tested in a 24-well plate by replica plating in 2 concentrations of 6TG (either 1 or 2 μM) as well as HAT (Figure 7-3–9). After selection, the cells were stained by methylene blue. MMR mutants should be resistant to 6TG and HAT, whereas *Hprt*-deficient mutants will be resistant to 6TG and sensitive to HAT. As expected, the results showed no *Hprt*-deficient mutants. Of the clones, 49 were strongly resistant to 6TG, 29 were weakly resistant and 25 were false positives since they were 6TG sensitive (Table 7-2). The weak resistance to 6TG may be due to a dose-dependent effect of heterozygous deletions of MMR genes.

Several selection controls were plated in parallel. The NGG5-3 cell line is the *Blm*-deficient parental cell line of the mutants with the genotype of $Blm^{-/-}Hprt^{+/+}$. With the presence of the *Hprt* gene, it was resistant to HAT but sensitive to 6TG (Figure 7-3, Figure 7-6). The NGG5-3 cell line was derived from the NN5 cell line, which was generated from the double targeted cell *Blm*(m1/m3) (Luo *et al.* 2000). NN5 was generated from the *Hprt*-deficient AB2.2 cell line and is thus $Blm^{-/-}Hprt^{-}$. Without the functional *Hprt* gene product, NN5 was sensitive to HAT but resistant to 6TG (Figure 7-8). The cell line ww56+*Hprt* ($Dnmt1^{-/-}Hprt^{+}$) has a double targeted *Dnmt1* gene and a heterozygously targeted *Gdf9* gene with an *Hprt* selection marker. As described by Guo, a *Dnmt1*-deficient cell was resistant to 6TG (Guo *et al.* 2004). With an *Hprt* gene, it was resistant to HAT, too (Figure 7-3, Figure 7-6). As expected, all control cell lines demonstrated the correct phenotype in hypoxanthine/aminopterin/thymine (HAT) or 6TG.

Table 7-2 6TG resistant mutant clones in the mutation library

A

Pool	PBA	PCA	PDA	PEA	PFA	PGA	PHA
Number of $6TG^R$ clones observed	7	20	12	13	20	12	19

B

Pool	**Phenotypes**			**Total number of clones**
	Strong	**Weak**	**Sensitive**	
PBA	2	3	2	7
PCA	7	6	7	20
PDA	9	2	1	12
PEA	6	4	3	13
PFA	15	5	0	20
PGA	4	2	6	12
PHA	6	7	6	19
Total number of clones	**49**	**29**	**25**	**103**

C

Clone ID	**Strong**	**Weak**	**Sensitive**	**Total number of clones**
B	6;7	1;2;3	4;5	7
C	1;2;4;5;10;15;17	3;11;13;14;18;19	6;7;8;9;12;16;20	20
D	1;2;3;4;5;7;8;9;12	6;11	10	12
E	1;3;4;5;9;14	2;7;11;13	6;8;12	13
F	1;2;3;4;6;8;11;12;13;15;16;17;18;19;20	5;7;9;10;14	0	20
G	6;8;9;10	1;4	2;3;5;7;11;12	12
H	1;3;5;6;13;14;	2;4;7;8;9;11;12;	10;15;16;17;18;19	19
Total number of clones	**49**	**29**	**25**	**103**

A. Number of clones isolated from each mutation pool; **B**. 6TG resistance phenotypes of primary clones; **C**. Scoring of 6TG resistance phenotypes of individual clones.

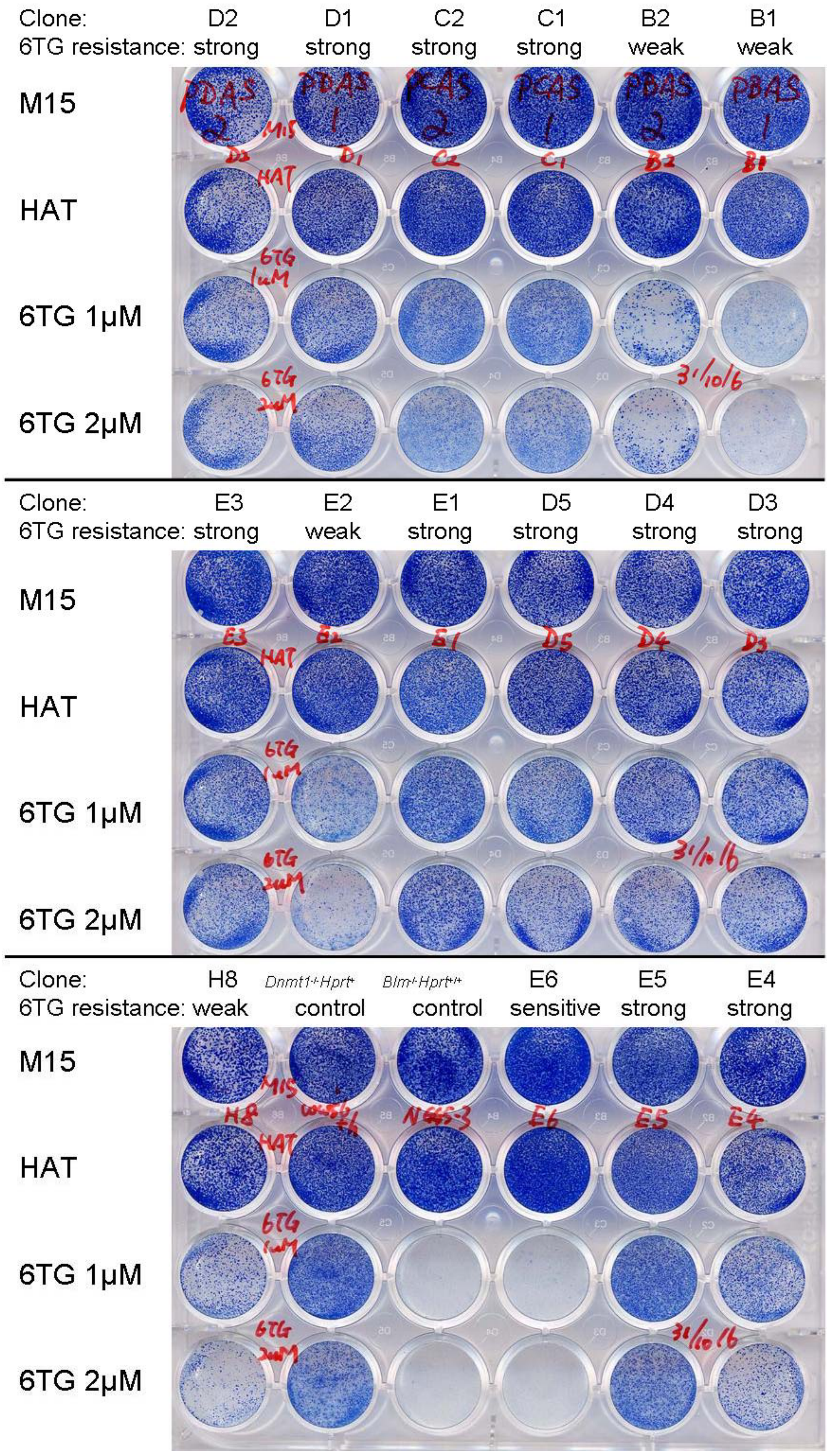

Figure 7-3 **Phenotype of mutant clones**

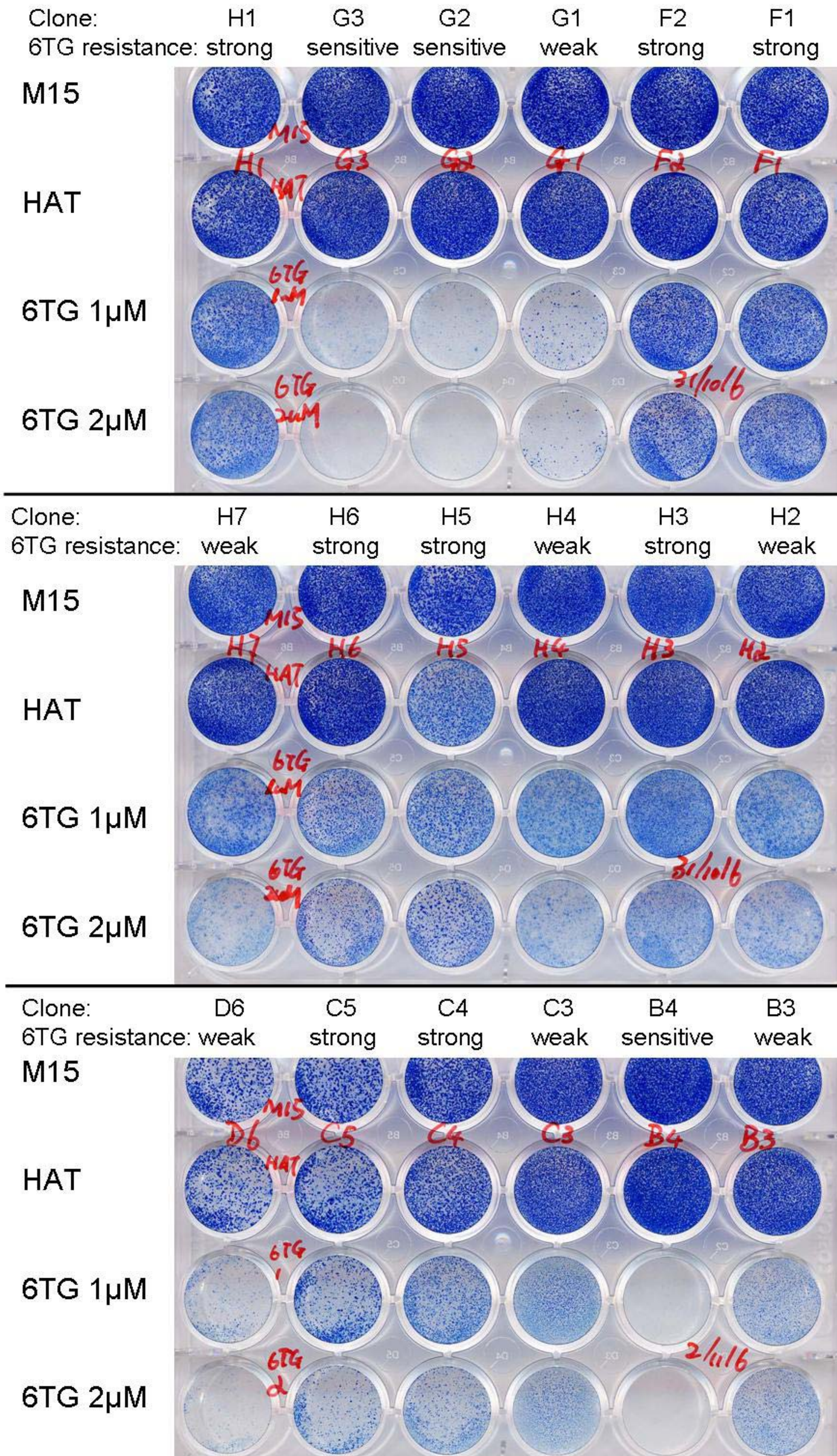

Figure 7-4 Phenotype of mutant clones

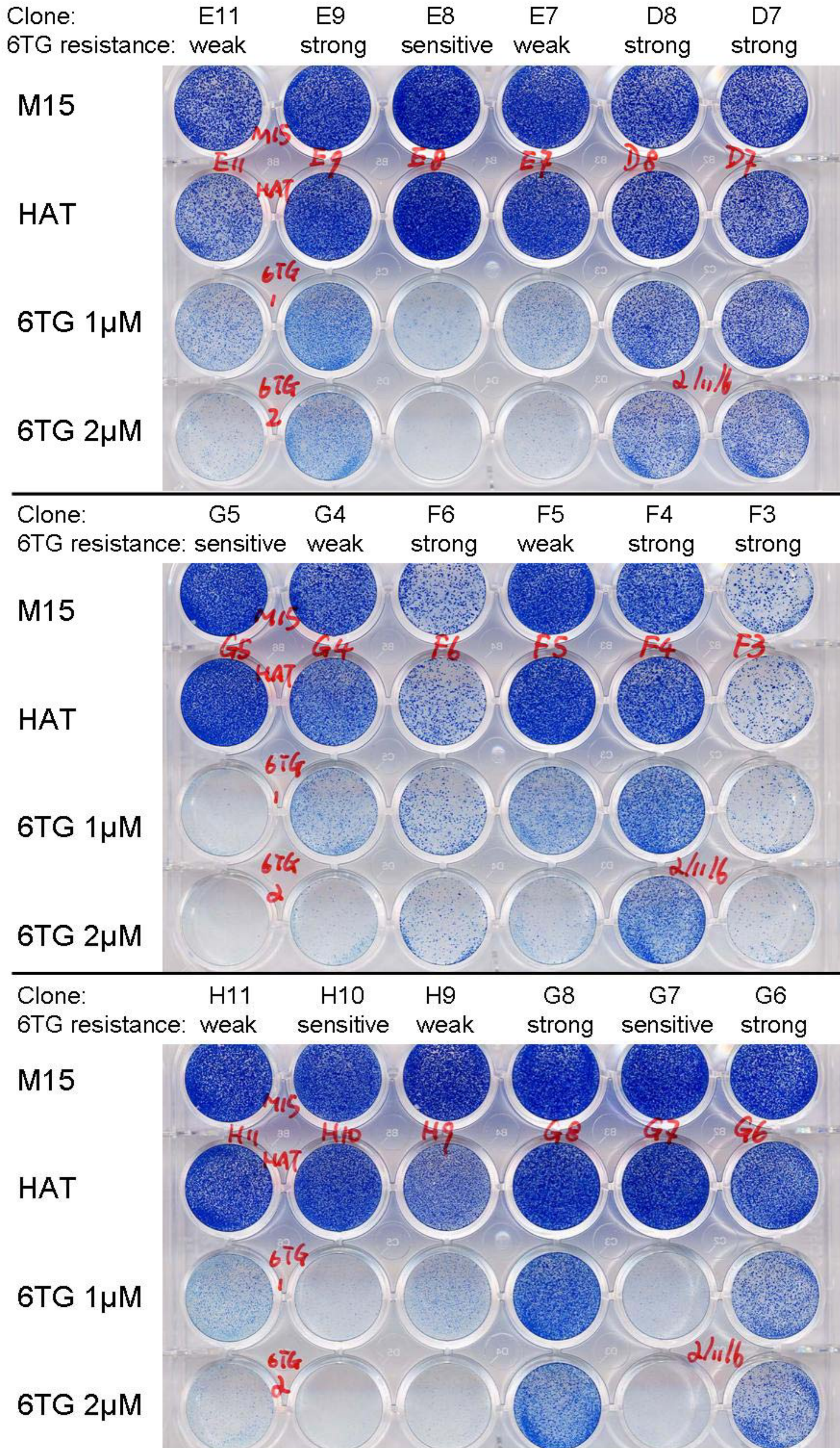

Figure 7-5 Phenotype of mutant clones

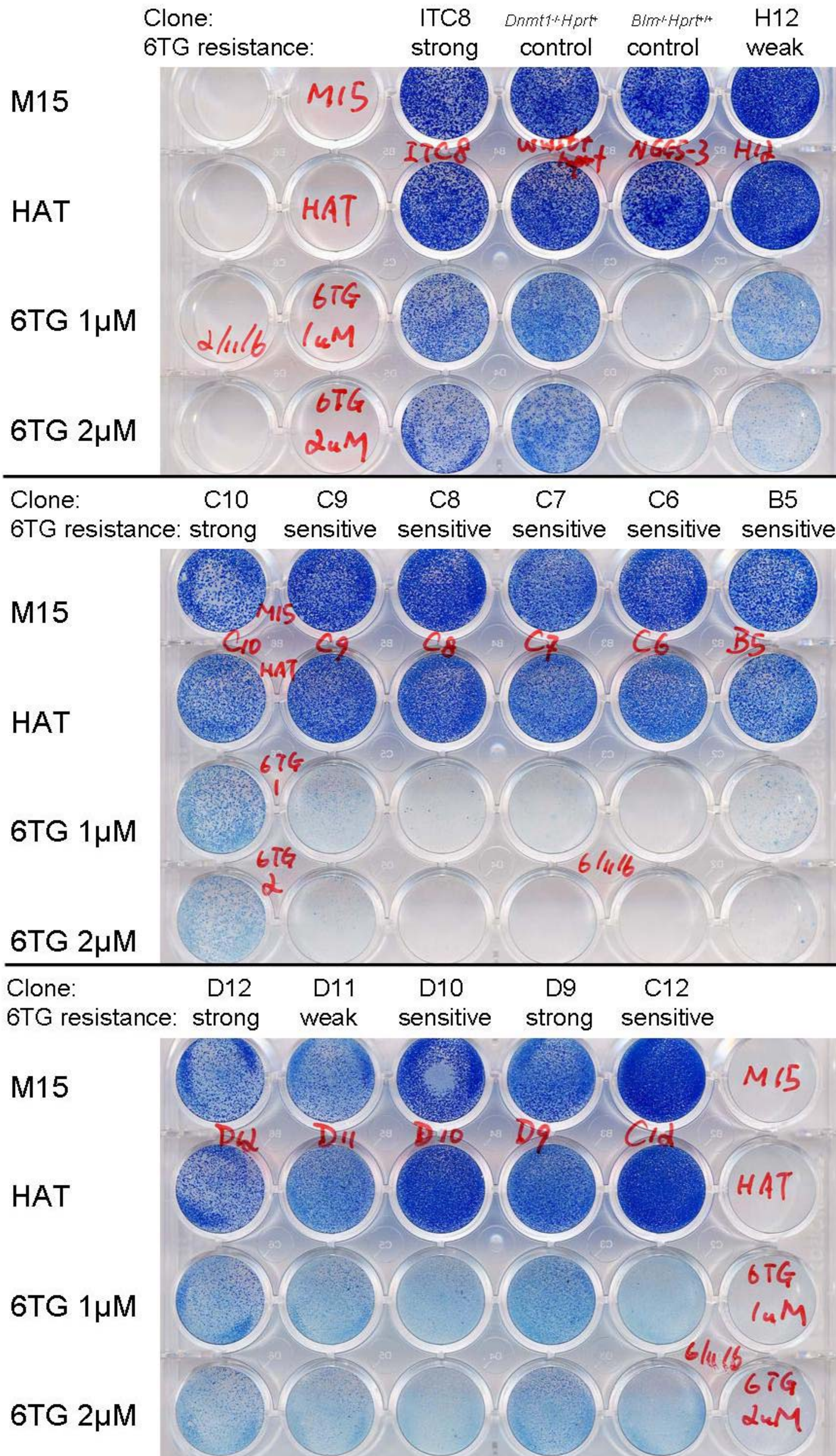

Figure 7-6 **Phenotype of mutant clones**

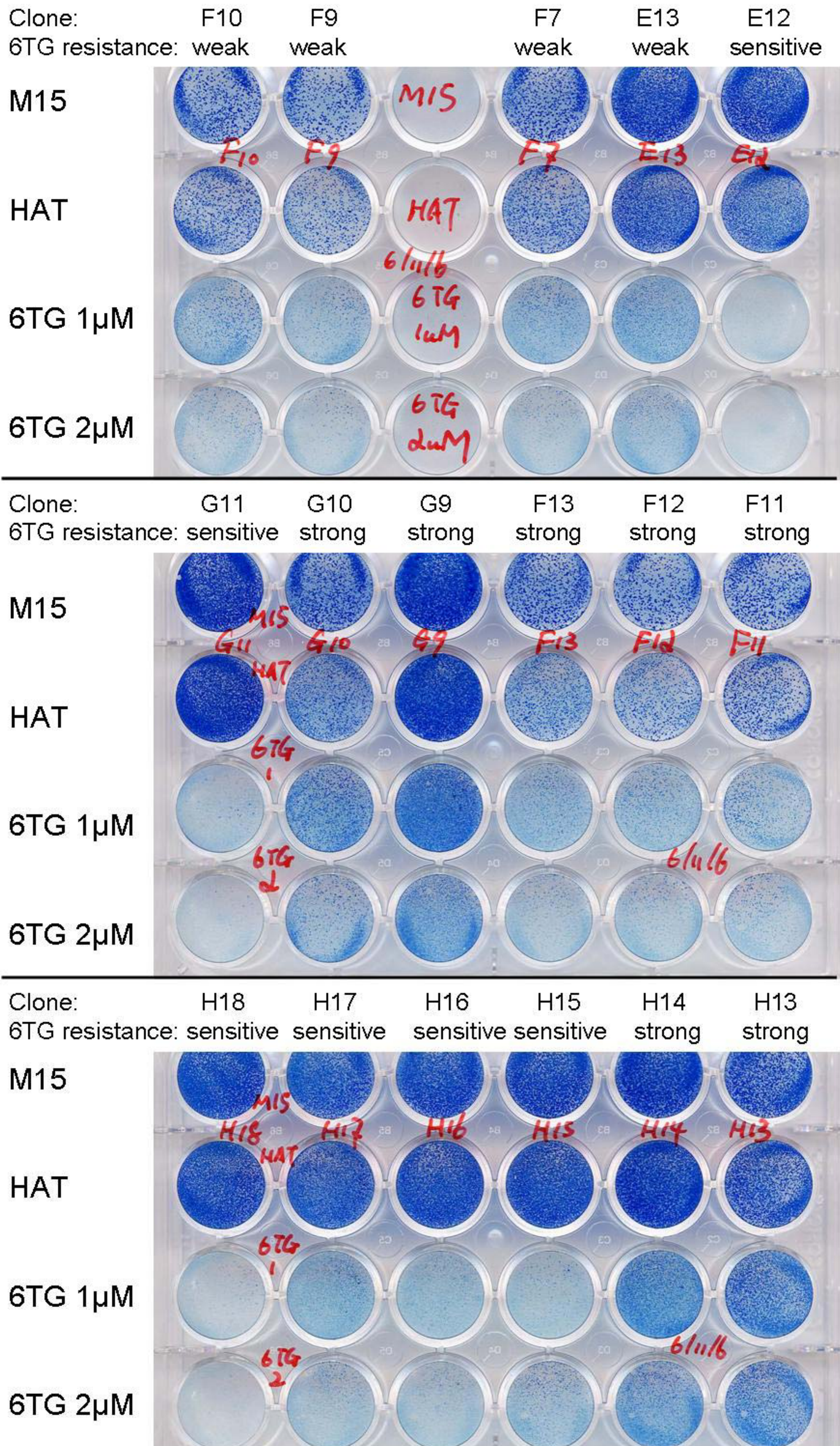

Figure 7-7 Phenotype of mutant clones

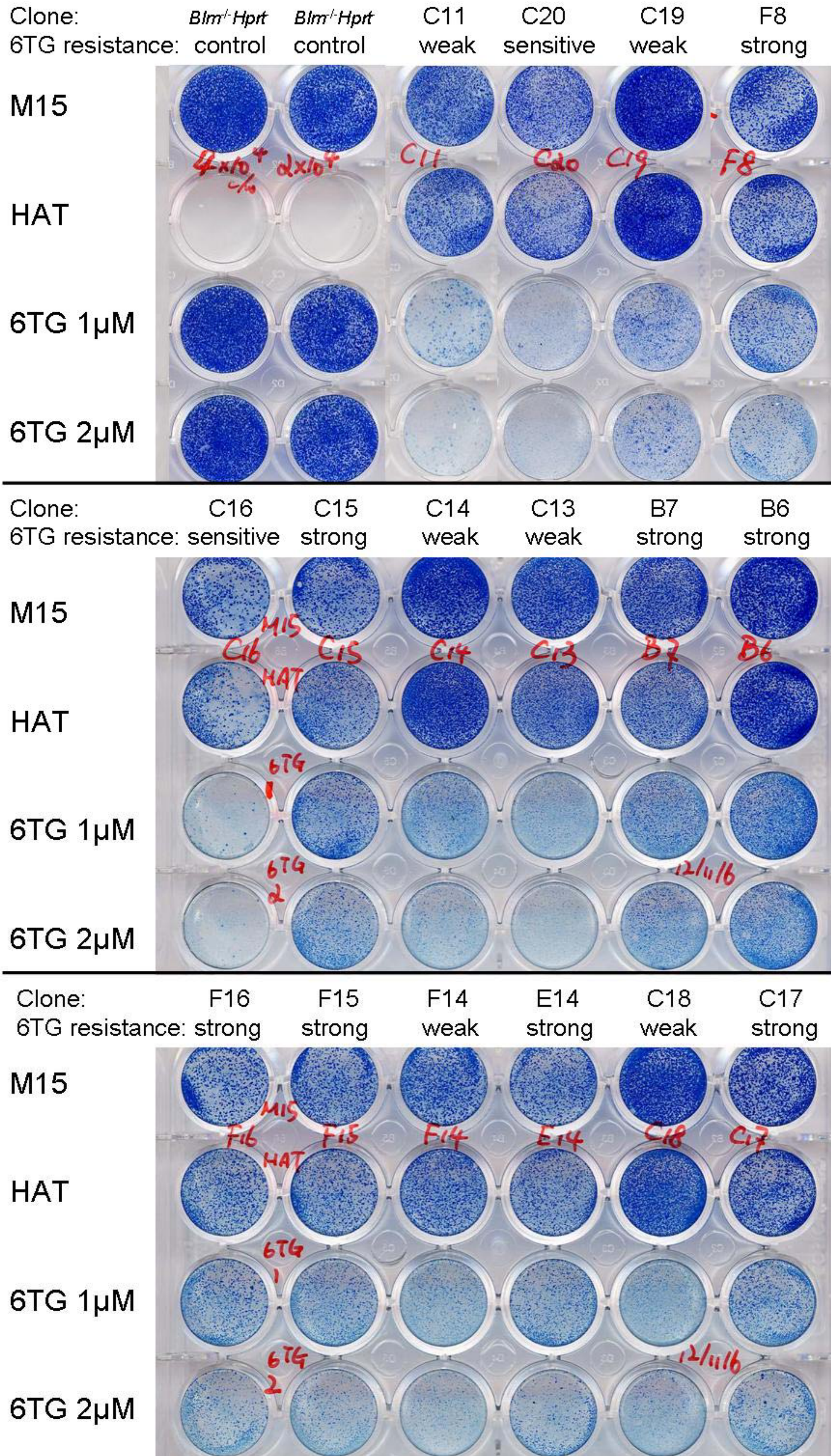

Figure 7-8 **Phenotype of mutant clones**

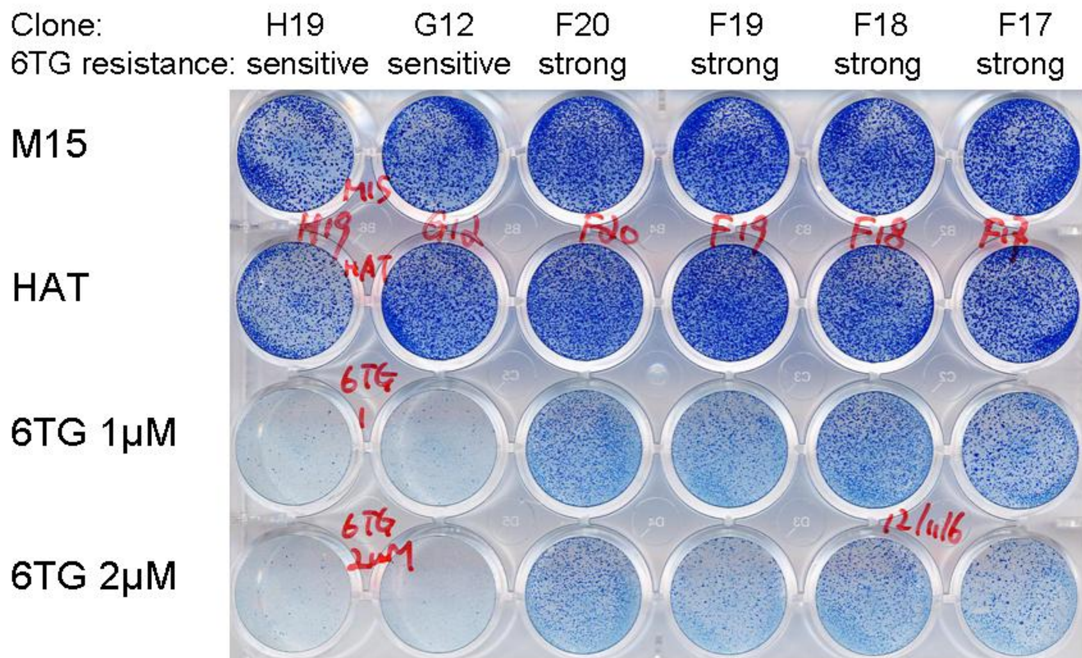

Figure 7-9 Phenotype of mutant clones

7.2.5 Clonal relationship of $6TG^R$ clones

The clonal relationships within a mutation pool were established by Southern blot analysis using a *PuroΔtk* probe (Figure 7-10). The *PuroΔtk* probe is a 1233 bp *Pst* I fragment from plasmid YTC37 (Chen, Y. T. *et al.* 2000) used to construct the retroviral vector. Genomic DNA from the clones was digested with *Hind* III, which cuts once in the provirus, thus each proviral insertion will generate a different-sized restrict fragment depending on the position of the *Hind* III site in the flanking genomic DNA. Independent clones will have different fragment sizes on the Southern blot. According to Southern blot analysis (Figure 7-11–13), 32 redundant mutants clones were identified. Twenty-seven unique MMR-deficient candidates were identified with strong resistance to 6TG (Table 7-3).

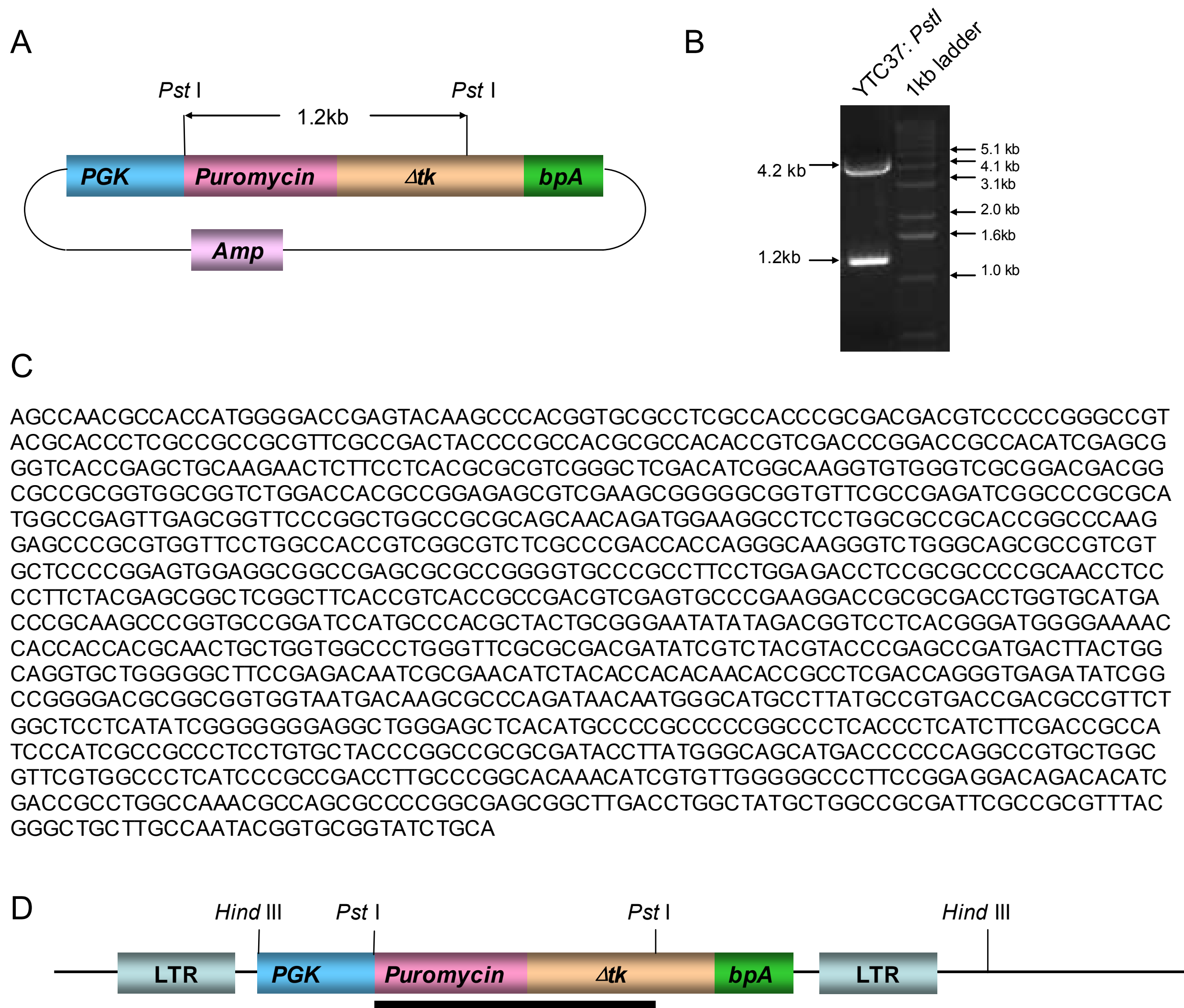

Figure 7-10 Design of the *PuroΔtk* probe and the Southern blot strategy to establish clonal relationships

A. Schematic view of *Pst* I sites in the *PuroΔtk* cassette. **B**. Isolation of the *PuroΔtk* probe. Plasmid YTC-37 was digested with *Pst* I, resulting in two fragments. The 1.2 kb fragment was purified for use as a Southern blot probe later. **C**. The sequence of the *PuroΔtk* probe from the DNA fragment between two *Pst* I sites of YTC-37 plasmid. **D**. Southern blot analysis strategy to establish clonal relationships between clones. A locus with the provirus and flanking genomic DNA is shown. By cutting genomic DNA of a mutant clone with *Hind* III, the DNA fragment containing the *PuroΔtk* cassette and the nearest *Hind* III site in the flanking DNA will hybridize with the *PuroΔtk* probe (black bar).

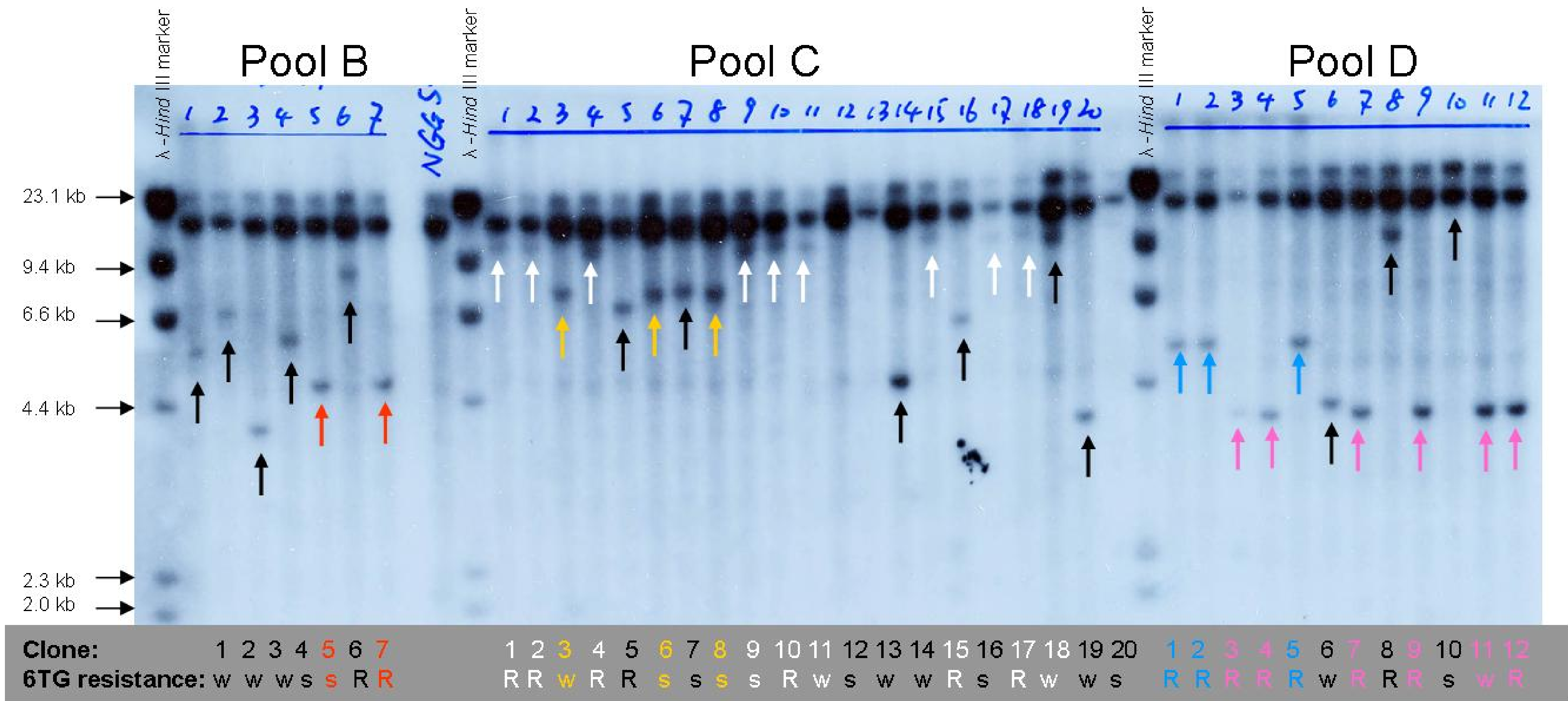

Figure 7-11 Southern blot to establish clonal relationships of $6TG^R$ clones in mutation pools B, C and D

Figure 7-11 Southern blot to establish clonal relationships of $6TG^R$ clones in mutation pools B, C and D

DNA of 7 mutants from pool B, 20 mutants from pool C and 12 mutants from pool D were digested with *Hind* III. There is a strong background band below the 23.1 kb marker. It exists in the control cell line (NGG5-3) and all mutants. When Blasting the *PuroΔtk* probe sequence in the mouse genome (NCBI m37, Ensembl release 47), there was no homologous sequence longer than 21 bp. Thus, this background band comes from shared vector sequences from the multiple targeting experiments used to construct NGG5-3. The junction fragments are indicated by arrows. The identical size of mutants' junction fragments within a mutation pool suggests clonal relationships and these fragments are indicated by arrows of a non-black colour. Different colours indicate different clonal relationships. Black arrows indicate clones without an obvious clonal relationship with others. Mutants C12 and C13 do not have obvious junction fragments. These two clones' junction fragments may be too close to the background band to be seen. The name of each mutant clone and its 6TG resistance is indicated at the bottom of the figure. (R: strongly 6TG-resistant; w: weakly resistant to 6TG; s: sensitive to 6TG.)

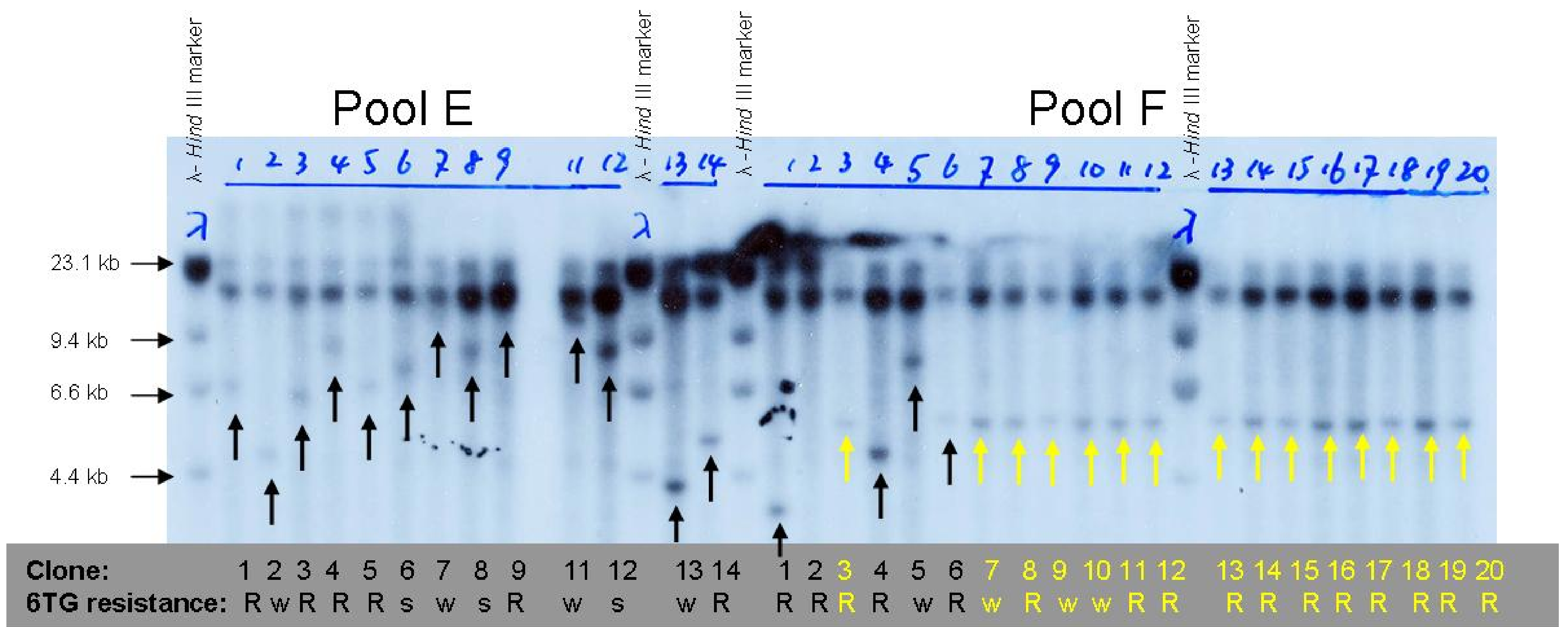

Figure 7-12 Southern blot to establish clonal relationships of 6TG[R] clones in mutation pools E and F

Figure 7-12 Southern blot to establish clonal relationships of $6TG^{R}$ clones in mutation pools E and F

DNA of 13 mutants from pool E and 20 mutants from pool F digested with *Hind* III and hybridized with the *PuroΔtk* probe. Black arrows indicate clones without an obvious clonal relationship with others. Yellow arrows indicate identical junction fragments in size. Mutant F2 does not have an obvious junction fragment. Its junction fragments may be too close to the background band to be observed. (R: strongly 6TG-resistant; w: weakly resistant to 6TG; s: sensitive to 6TG.)

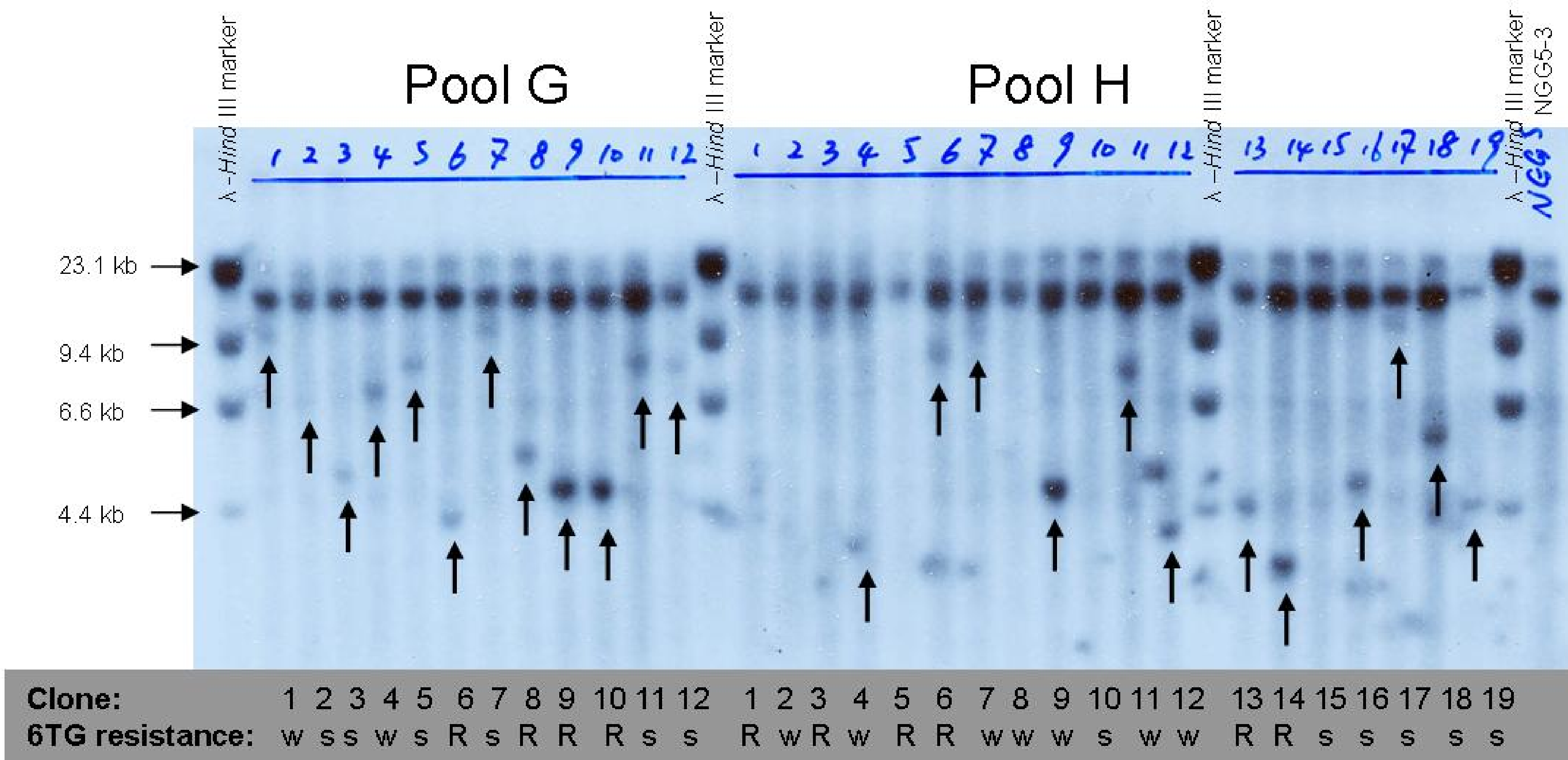

Figure 7-13 **Southern blot to establish clonal relationships of $6TG^R$ clones in mutation pools G and H**

Figure 7-13 Southern blot to establish clonal relationships of $6TG^R$ clones in mutation pools G and H

DNA of 12 mutants from pool G and 19 mutants from pool H was digested with *Hind* III and hybridized with the *PuroΔtk* probe. The junction fragments are indicated by arrows. There are no junction fragments that are identical in size. Mutants H1–3, 5, 8, 10 and 15 do not have obvious junction fragments. Their junction fragments may be too close to the background band to be seen. (R: strongly 6TG-resistant; w: weakly resistant to 6TG; s: sensitive to 6TG.)

Table 7-3 **Redundant mutant clones and unique mutant clones**

Pool	Number of unique clones	ID of unique clones	Number of redundant clones	ID of redundant clones	Number of unique clones with strong 6TG resistance	ID of unique clones with strong 6TG resistance
B	6	1–4, 6–7	1	5	2	6, 7
C	8	1, 3, 5, 7, 14, 16, 19–20	10	2, 4, 6, 8–11, 15, 17, 18	2	1, 5
D	5	1, 4, 6, 8, 10	7	2–3, 5, 7, 9, 11–12	3	1, 4, 8
E	13	1–13	0	None	5	1, 3, 4, 5, 9
F	6	1, 2, 4–6, 16	14	3, 7–15, 17–20	5	1, 2, 4, 6, 16
G	12	1–12	0	None	4	6, 8, 9, 10
H	19	1–19	0	None	6	1, 3, 5, 6, 13, 14
Total number	69	69	32	32	27	27

In pools B–H, redundant mutant clones were identified by the clonal relationship established through Southern blot analysis. One clone of a number of redundant clones is chosen to be a representative to be analysed. These representatives and unique clones are chosen for array CGH and expression array analysis.

7.2.6 Array CGH analysis

A total number of thirty-four mutant clones were analysed by array CGH to identify copy number changes of genomic DNA. Twenty-seven are strongly 6TG resistant unique clones (Table 7-3) and seven are weakly 6TG resistant clones (B2, B3, C3, D6, E2, F9 and G4). The reference DNA for these experiments was genomic DNA from NGG5-3, extracted from the same passage of cells used for the virus tagging experiment. This is important because cells accumulate mutations during cell culture. The arrays used for this experiment were mouse tiling BAC clones with a total 18,294 clones. A dye swap experiment or two duplicate array experiments were preformed on each pair of mutant DNA and reference DNA. Each data point is derived from either a dye swap (two data) or two duplicates of a given BAC. Due to the high cost of array CGH, no more replicating experiments were performed. To obtain more confident data, adjacent BACs were used to estimate range of variation and to excluded low quality data points. An average Log_2(ratio) was calculated for a continuous region including BACs with similar Log_2(ratio). Within an experiment, the Log_2(ratio) between -0.29 and +0.29 for the signal from any BAC probe was regarded as no copy number change. The position of a BAC probe on the array is recorded as the position of its middle point (in the base pair of the chromosome). This method is not accurate to base pairs resolution because we cannot identify which part of a BAC sequence is contributing signal on the array. Thus, we have to bear in mind that even if two cell lines have the same BAC(s) deleted, they will not have the same starting and ending base pairs.

7.2.6.1 *Msh2* and *Msh6* homozygous deletions

As previously described, cells deficient for the MMR genes *Msh2, Msh6, Mlh1* or *Pms2* are resistant to 6TG (Abuin *et al.* 2000; Karran *et al.* 1996). In addition, *Dnmt1* deficiency gives 6TG resistance (Guo *et al.* 2004). Therefore, homozygous deletions of these genes were expected in the mutants. Of the thirty-four clones (27 strongly $6TG^R$, 7 weakly $6TG^R$) analysed, six clones (B6, D1, D4, D8, F4 and H14) were identified with homozygous deletions on chromosome 17 between 87.3 Mb and 91.7 Mb (Table 7-4). These deletions covered the genes *Msh2* and *Msh6*. Five clones (B6, D4, D8, F4 and H14) deleted both genes while one (D1) deleted *Msh6* only (Figure 7-14). In addition, nine clones (C1, C3, C5, D6, F9, F16, G8, G9 and H13) contain heterozygous deletions covering both *Msh2* and *Msh6*.

The six mutants with homozygous deletions covering *Msh2* and *Msh6* genes are from four mutation pools and they are not clonal according to the previous data of clonal relationships

built by Southern blot analysis. These deletions range from 2.6 to 4.4 Mb (Table 7-4). The average size of a homozygous deletion is 3.6 Mb. The Log_2(ratio) ranges from -2.21 to -1.16 and the average Log_2(ratio) is -1.67. The array CGH profiles of two of these mutants (D1 and D4) are shown in Figure 7-15 and Figure 7-16.

Mutant clone D1 was identified in which the *Msh6* gene was homozygously deleted without affecting the *Msh2* gene (Figure 7-15). This clone D1, from the mutation pool D, is a fully 6TG resistant mutant. It covers a 3.3 Mb deletion and the average Log_2(ratio) of the BACs in this region is -1.61. In addition to this homozygous deletion, other deletions are observed on chromosomes 12 (14 Mb) and 16 (1.0 Mb), and duplications are observed on chromosomes 10 (2.1 Mb), 13 (51 Mb) and 17 (1.8 Mb, 2.3 Mb). If *Msh6* was not a known MMR gene, it would be easy to identify which deleted gene in this clone results in the 6TG resistance, because the other regions are heterozygous.

In the Figure 7-16, the Log_2 deviation on chromosome 17 indicates a homozygous deletion including *Msh2* and *Msh6* genes. This clone (D4) has also lost chromosome Y and duplicated half of chromosome 1. The single BAC homozygous deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are observed in many samples, these are associated with chromosome Y loss. Thus, they were assumed to belong to chromosome Y but were incorrectly allocated to other chromosomes.

The tile path BACs at the *Msh2* and *Msh6* loci are shown in Figure 7-17. The zoomed-in figure of *Msh2* and *Msh6* homozygous deletions are shown in Figure 7-18 and the schematic view of these deletions is shown in Figure 7-19. This figure illustrates that the minimal deletion of these clones includes both the *Msh2* and *Msh6* genes, and the minimal common homozygous deletion (2 Mb) of the six clones. Thus, in the absence of any knowledge about these genes I would have identified them by my strategy. These clones confirmed the screen strategy and the capacity to identify deletions generated by irradiation through array CGH technology, it also confirmed the ability of the *Blm*-deficient system to generate homozygous deletions. It is noticed that all homozygous deletions do not cover the region more distal than 91.7 Mb. This is probably because the deficiency of the *Adcyap1* gene (93.6 Mb) may result in cell lethality, implied by high postnatal mortality in its knockout mice (Gray *et al.* 2001).

Table 7-4 Summary of homozygous deletions covering *Msh2* and/or *Msh6*

Clone	Deleted region starting BAC (Mb)	Deleted region ending BAC (Mb)	Deletion length (Mb)	Average Log_2(ratio) over deleted region
B6	RP24-315D17 (88.1)	RP23-340K6 (91.7)	3.6	-1.16
D1	RP23-476A24 (88.4)	RP23-340K6 (91.7)	3.3	-1.61
D4	RP24-343P24 (87.6)	RP23-340K6 (91.7)	4.1	-1.58
D8	RP23-10B20 (87.8)	RP24-82E21 (90.4)	2.6	-2.21
F4	RP23-243O16 (87.9)	RP23-340K6 (91.7)	3.8	-2.09
H14	RP24-107B10 (87.3)	RP24-494L17 (91.7)	4.4	-1.36

Five clones (B6, D4, D8, F4 and H14) deleted the both *Msh2* and *Msh6* genes while one (D1, blue font) deleted *Msh6* only. The sized, starting point, ending point and the average Log_2(ratio) over deleted region are shown here. These homozygous deletions may start or end at the same BAC(s), thus the genomic position of the BACs can be the same. However, the position of a BAC is annotated by the middle point position of this BAC, and a BAC probe does not have to be hybridized in full size to contain a signal. Therefore, the genomic positions of the starting and ending BACs of a deletion can not indicate the actual starting and ending points of a deletion. Genomic position from Ensembl release 48 (NCBI m37).

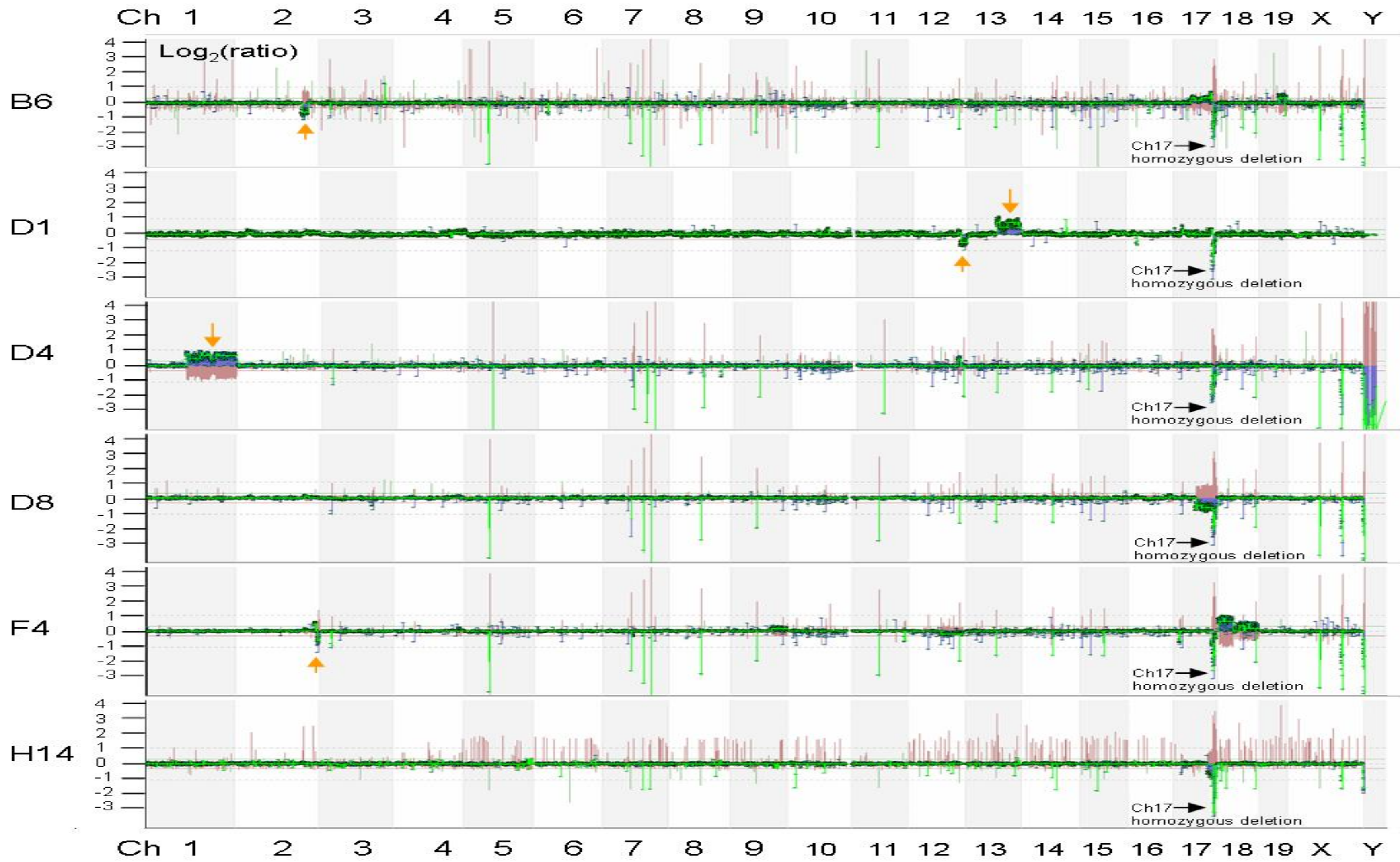

Figure 7-14 ***Msh2* and *Msh6* homozygous mutations**

Figure 7-14 *Msh2* and *Msh6* homozygous mutations

Six mutant clones with strong 6TG resistance were analysed by array CGH. Chromosomes were also placed horizontally in ascending order followed by chromosomes X and Y. A dot indicates the middle point position of a given BAC. The Y axis is the Log_2(Mutant DNA signal/Reference DNA signal) of the BAC probes. This Log_2(ratio) indicates a duplication or amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log_2(Reference DNA signal/Mutant DNA signal) of BAC clone probes from the reciprocal hybridization experiment. Besides the homozygous deletion covering *Msh2* and *Msh6*, there are other genomic DNA copy number changes. For example, clone B6 has a heterozygous deletion on chromosome 2 (yellow arrow), clone D1 has a duplication on chromosome 13 and a heterozygous deletion on chromosome 12 (yellow arrow), clone D4 has distal half of chromosome 1 duplicated (yellow arrow), and clone F4 has a heterozygous deletion on chromosome 2 (yellow arrow). Some small homozygous deletions (single BAC size or two-BAC size) on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are visible in the clones B6, D4, D8 and F4. A common change for these four clones is that they lost all or a big part of chromosome Y. It is assumed that these small deletions should belong to chromosome Y but were incorrectly annotated to other chromosomes.

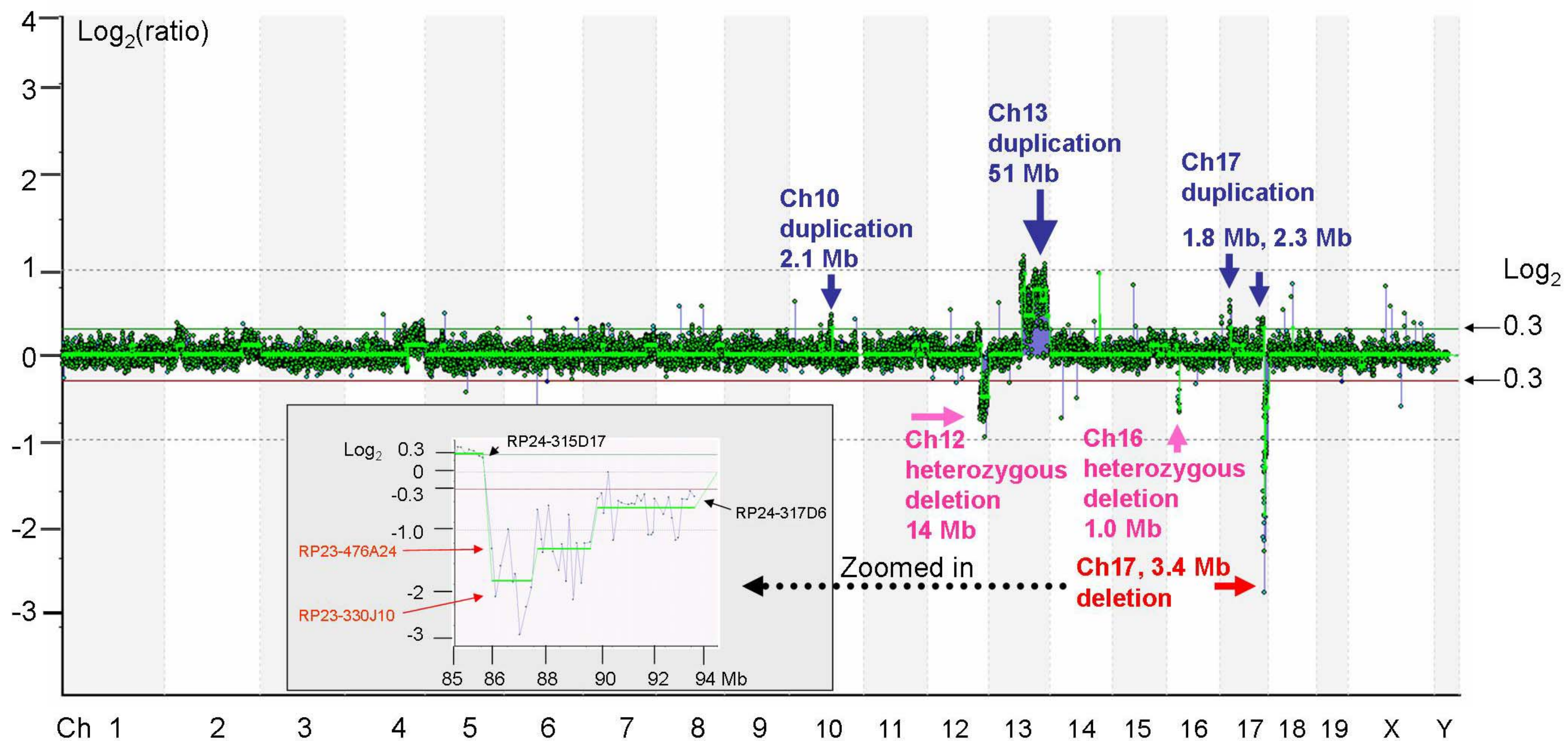

Figure 7-15 **Array CGH profile of mutant D1 with a homozygous deletion at *Msh6***

Figure 7-15 Array CGH profile of mutant D1 with a homozygous deletion at *Msh6*

The Log_2(ratio) of mutant:*Blm*-deficient cell DNA was calculated for each BAC and plotted along each mouse chromosome. Chromosomes were also placed in ascending order followed by X and Y. Each dot indicates a given BAC's middle point position. The Y axis values the Log_2(Mutant DNA signal/Reference DNA signal) of a given BAC probe. The Log_2(ratio) indicates an amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of the BAC clone probes. A red arrow indicates a homozygous deletion including gene *Msh6* on chromosome 17. A zoomed-in picture shows the details of this deletion. BAC RP23-330J10 and RP23-476A24 (red font) are in a homozygously deleted region which includes *Msh6*. BAC RP24-315D17 (containing gene *Msh2*) is duplicated with Log_2(ratio) of ~0.3. BAC RP24-317D6 is the most distal BAC on chromosome 17 in the 200 kb resolution array CGH. In addition, there are deletions (pink arrows) on chromosomes 12 and 16 and duplications (blue arrows) on chromosomes 10, 13, and 17. Some BACs with blue lines are regarded as noise signals because these signals cannot be replicated. The grey horizontal dotted lines are for Log_2(ratio) references of either +1 or -1.

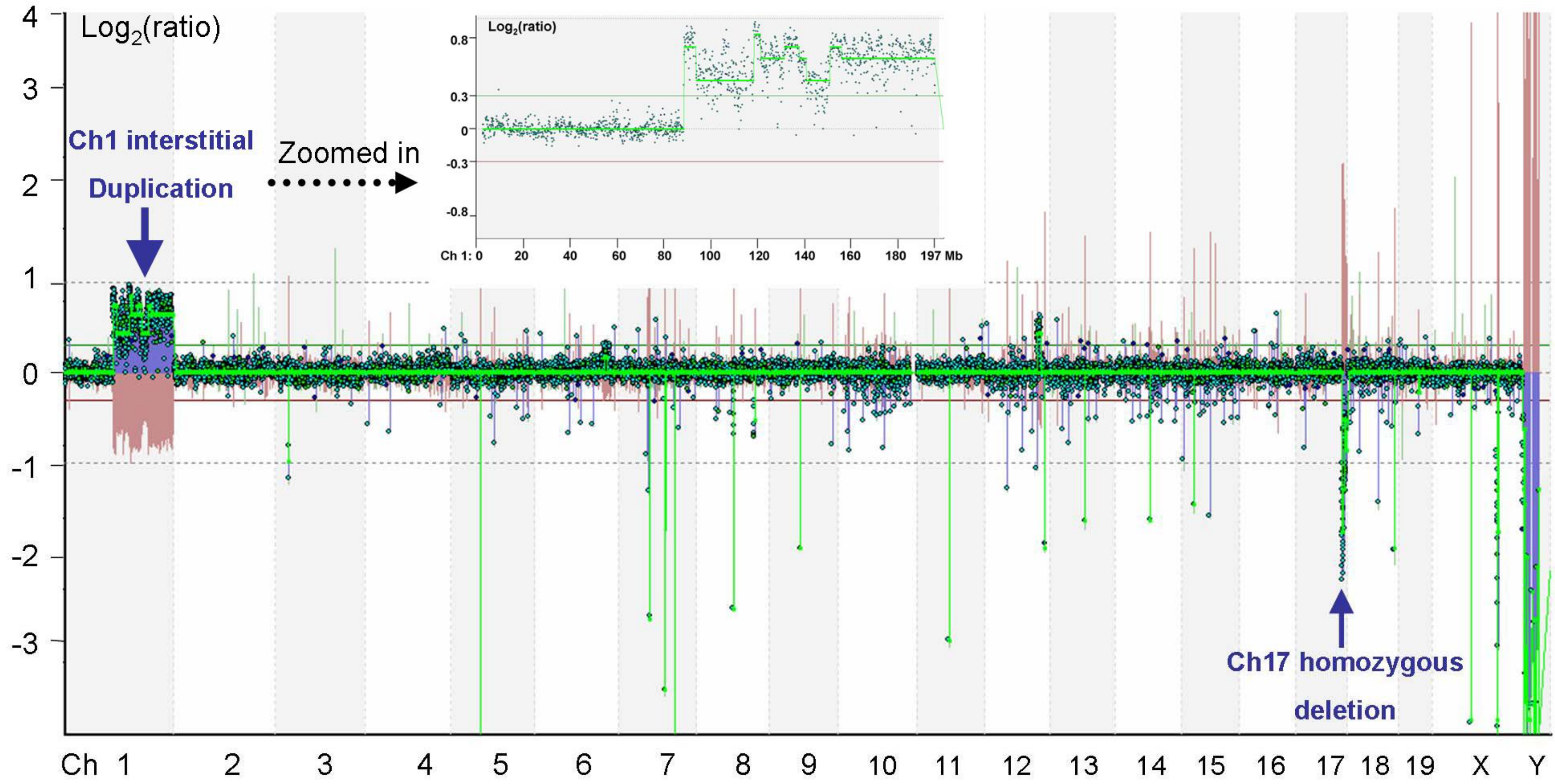

Figure 7-16 **Array CGH analysis of mutant D4**

Figure 7-16 Array CGH analysis of mutant D4

Mutant clone D4 is presented to illustrate how data from array CGH analysis was interpreted. BACs are aligned on the X axis in ascending order according to their genomic position in every chromosome. Chromosomes are also placed in ascending order followed by X and Y. A dot indicates the middle point position of a given BAC. The Y axis is the Log_2(Mutant DNA signal/Reference DNA signal) of a given BAC probe. This Log_2(ratio) indicates a duplication or amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log_2(Reference DNA signal/Mutant DNA signal) of BAC clone probes from the reciprocal hybridization experiment. The green line below to -2 on chromosome 17 indicates a homozygous deletion including genes *Msh2* and *Msh6*. A zoomed-in picture shows half of chromosome 1 is duplicated. This clone has lost the Y chromosome and the distal half of chromosome 1 is duplicated. There are some single BAC deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X. In many experiments, these deletions were found to be associated with chromosome Y loss. Thus, they were assumed to belong to chromosome Y but were incorrectly annotated to other chromosomes. The grey dotted lines are for Log_2(ratio) references of either +1 or -1.

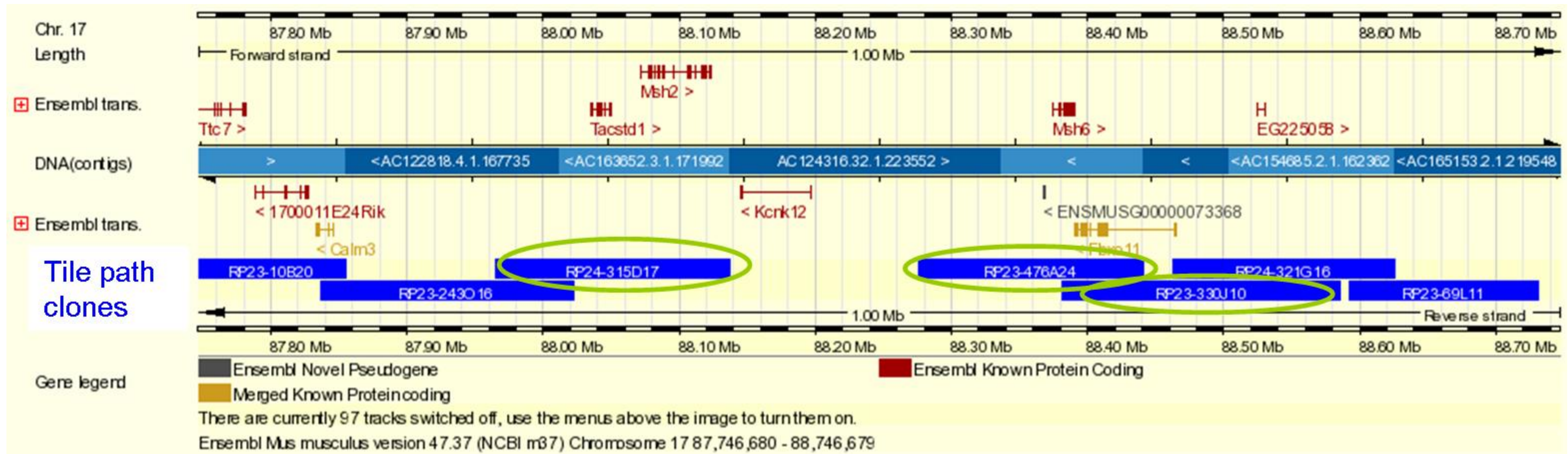

Figure 7-17 Tile path BACs at *Msh2* and *Msh6* loci

A snapshot of the 1.0 Mb chromosome region between 87,746,680–88,746,679 bp on mouse chromosome 17 is shown here (NCBI m37, Ensembl release 47). The Ensembl known protein coding sequences are shown in red. Blue boxes indicate BACs used on the tile path array CGH (200 kb resolution). BAC RP24-315D17 (one of the green circles) covers gene *Msh2*. BAC RP23-476A24 and RP23-330J10 (the other two green circles) cover gene *Msh6*. There is a gap between BAC RP24-315D17 and RP23-476A24.

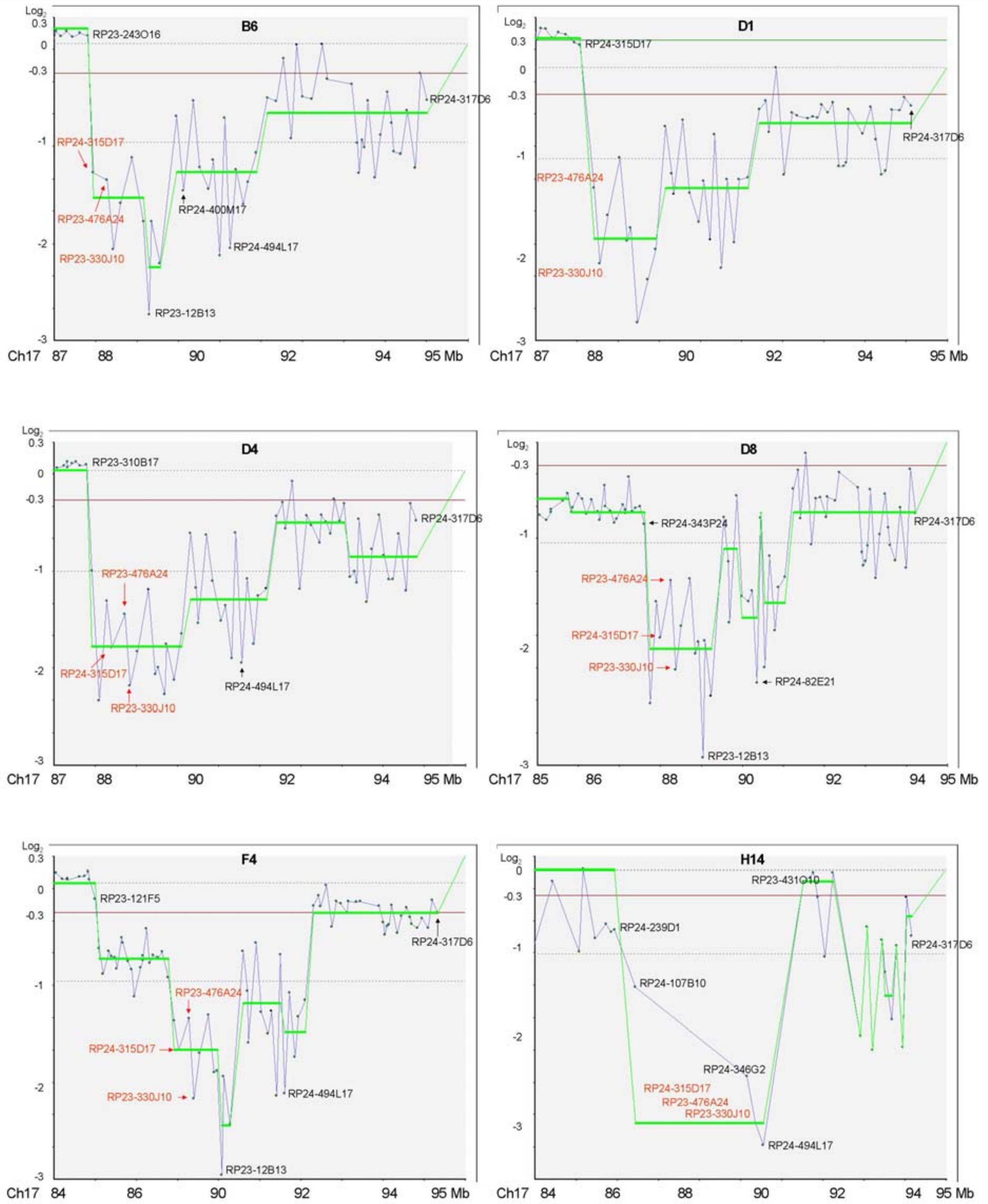

Figure 7-18 Zoomed-in regions of *Msh2* and *Msh6* homozygous deletions

Figure 7-18 Zoomed-in regions of *Msh2* and *Msh6* homozygous deletions

Array CGH profiles of the mutant clones B6, D1, D4, D8, F4 and H14 are shown on part of chromosome 17. The Log_2(Mutant DNA signal/Reference DNA signal) is calculated for each BAC and plotted against their genomic positions. A blue line connects the nearest BACs. A light green line indicates the average Log_2(Mutant DNA signal/Reference DNA signal) of regions. BAC RP24-315D17 (red font except in D1) covers the gene *Msh2*, while BACs RP23-476A24 and RP23-330J10 (red font) represent a region covering the gene *Msh6*. Log_2(ratio) of these three BACs is below -1 (except RP24-315D17 in D1) therefore the two genes are regarded as homozygously deleted. However, the two homologous chromosomes are not homozygous. Some BACs nearby are annotated in a black font. BAC RP24-317D6 is the most distal BAC of chromosome 17 on the 200 kb resolution array CGH.

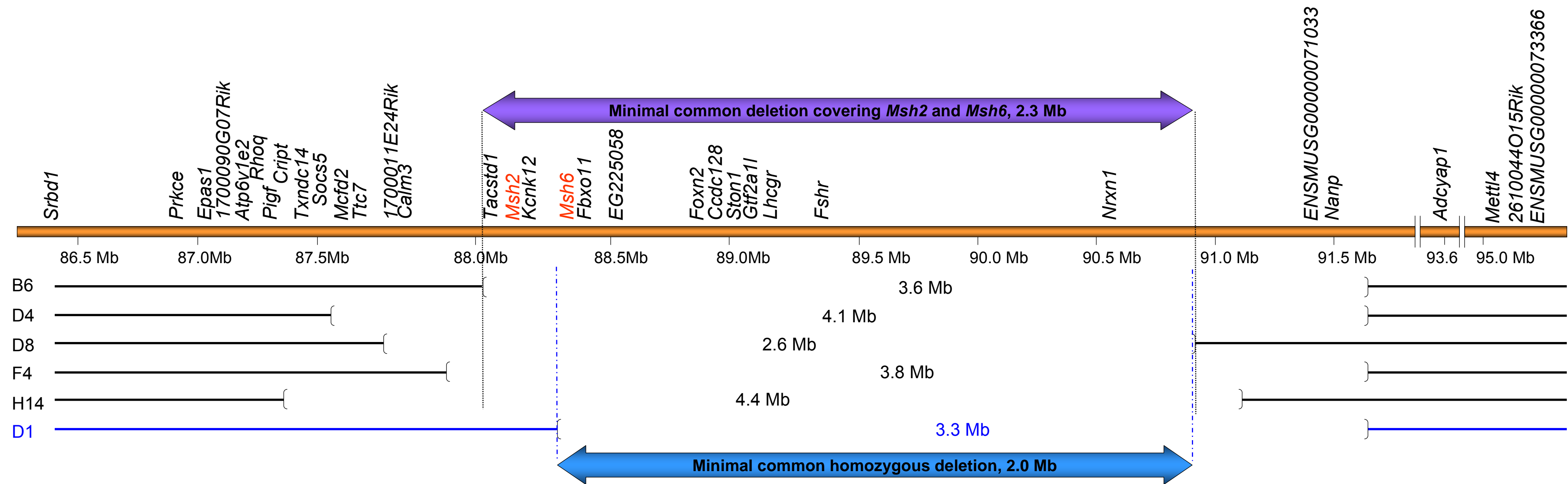

Figure 7-19 Schematic view of *Msh2* and *Msh6* homozygous deletions in six mutants

The region of 86.5–95.1 Mb of mouse chromosome 17 is shown with genes annotated. *Msh2* and *Msh6* loci are shown in a red font. Genomic DNA of the mutant clones B6, D4, D8, F4, and H14 is shown in black lines and genomic DNA of the mutant D1 is shown in blue lines. Homozygous deletions with their sizes in this region are shown in brackets. The first five mutant clones contain homozygous deletions covering *Msh2* and *Msh6* genes. The deletion in clone D1 does not contain *Msh2*. The minimal common deleted region in the first five mutants is 2.3 Mb in size and the minimal common homozygous deletion in these six clones is 2.0 Mb in size.

7.2.6.2 *Msh2* and *Msh6* heterozygous deletions

In addition to the six clones with homozygous deletions, nine clones were identified which had heterozygous deletions of the same region on chromosome 17. These deletions are believed to be heterozygous because the Log_2(Mutant DNA signal/Reference DNA signal) is higher than -1 and below zero. Their Log_2(ratio) is actually between -0.4 and -0.6 (Figure 7-20). The nine mutants are C1, C3, C5, D6, F9, F16, G8, G9 and H13. Their array CGH profiles are shown in Figure 7-20 for the whole genome and in Figure 7-21 for chromosome 17. These mutants are isolated from five mutation pools (C, D, F, G and H). Six of them (C1, C5, F16, G8, G9 and H13) are strongly resistant to 6TG. Three of them (C3, D6 and F9) are weakly resistant to 6TG. Among these mutants, the retroviral tagging revealed that F9 and F16 are derived from the same clone thus their array profiles are identical. This validates the reliability of the array CGH method. Within fifteen deletions and duplications in these two mutants, half of them are the same and the other half deletions and duplications have Log_2(ratio) which is close to either -0.3 or 0.3, the threshold set up for the BlueFuse software. Therefore small fluctuation of Log_2(ratio) resulted in different call of regions. However, this can be calibrated using manual examination.

These nine deletions range from 13.3 Mb to 46 Mb in size (Table 7-5), much larger than the sizes of homozygous deletions covering the *Msh2* and *Msh6* genes (Table 7-4). The average length of heterozygous deletion is 28 Mb. Except one clone, G9, deletions in the eight clones extend to the last BAC probe on chromosome 17, RP24-317D6, suggesting that these eight clones lost the whole distal part of chromosome 17. These clones also have other duplications and deletions in their genome. Although these mutant clones still have one copy of genes *Msh2* and *Msh6* in the genome, six of them have shown strong 6TG resistance. It is assumed that the undeleted allele of either *Msh2* or *Msh6* has a small mutation, which results in a loss-of-function of either *Msh2* or *Msh6.* These mutations were not identified by 200 kb resolution array CGH probably due to their small size. It is anticipated that this hypothesis will be confirmed by analysis of *Msh2* and *Msh6* transcripts.

Monosomies and trisomies were found in these mutants. For example, clone D6 appears to have a monosomy of chromosomes 2, 9, 13 and 16 (Figure 7-22), while clone G9 appears to have trisomies of chromosome 6 and 18 (Figure 7-20). In the clones C5, F9 and F16, some small homozygous deletions (single BAC) deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are associated with chromosome Y loss. This also confirmed the assumption that these BACs should be annotated to chromosome Y. The

clone D6 was selected to show chromosome copy number changes (Figure 7-22). This clone contains a 5.8 Mb heterozygous deletion including *Msh2* and *Msh6* genes on chromosome 17. A deletion on chromosome 2 may be homozygous. In addition, there seem to be monosomies on chromosomes 2, 9, 13 and 16 and trisomies on chromosome 1 and 14. Chromosome 11 appears to contain a segmental deletion at the centromere end. Several facts indicate that this clone might be a mosaic population or a tetraploid cell or both. Firstly, the Log_2(ratio) of chromosome Y is near -0.3. The ES cells I used are XY cells. If chromosome Y is lost, the Log_2(ratio) of chromosome Y on the array CGH should be like those of homozygous deletions. If one chromosome Y remains in a tetraploid cell, the Log_2(ratio) of it should be close to that of heterozygous deletions. However, the Log_2(ratio) of chromosome Y in clone D6 is near 0.3, suggesting a impure clone. Secondly, monosomies are very rare in ES cells. Thus, the clone might be a mosaic population of normal ploidy cell with tetraploid cells with one copy of chromosome loss (Log_2(¾)= -0.42). Thirdly, the Log_2(ratio) of chromosome 1 and 14 is less than 0.32, the Log_2 of a pentaploid chromosome in a tetraploid cell (Log_2(5/4)), indicating this clone is mosaic.

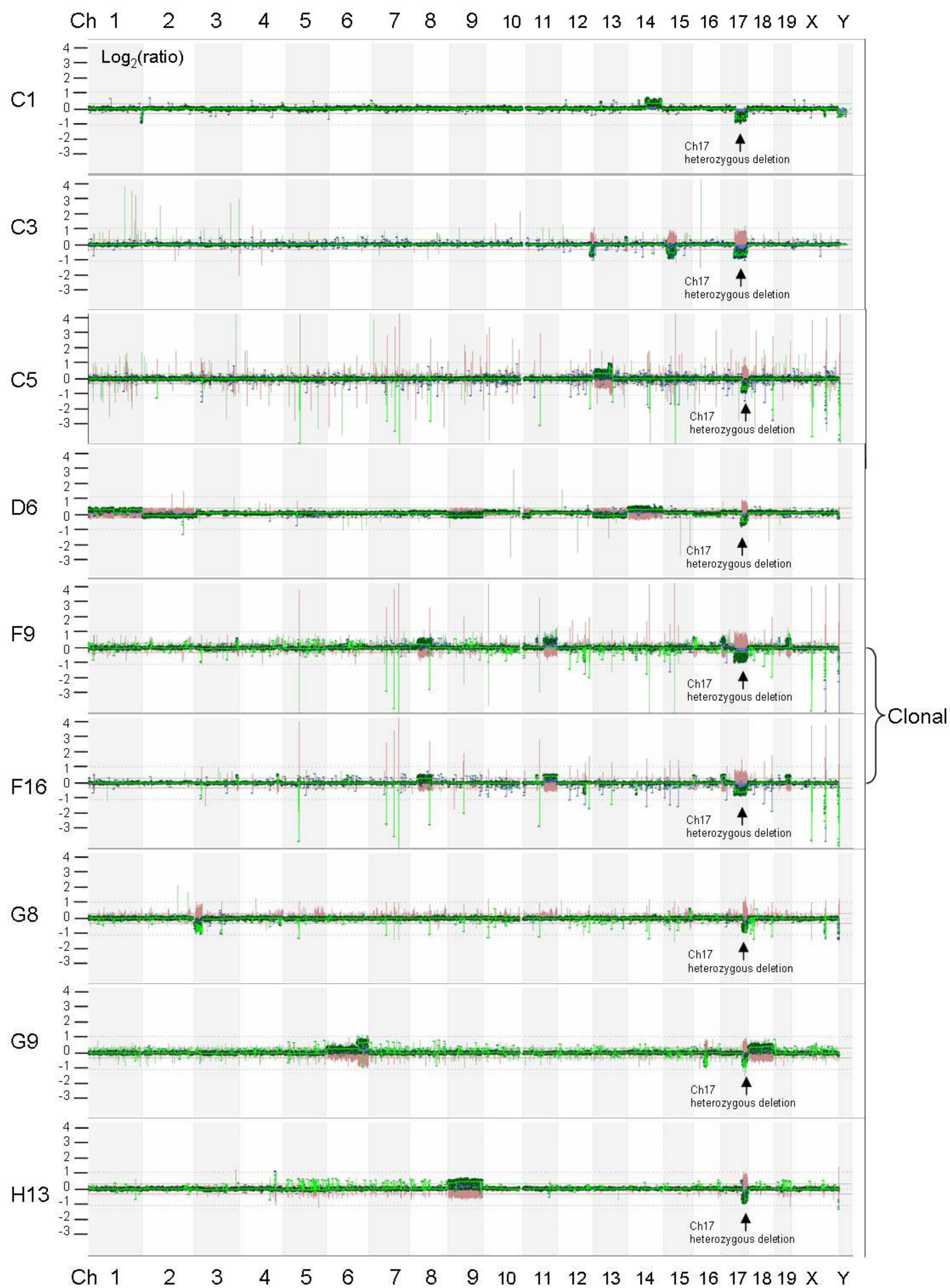

Figure 7-20 ***Msh2* and *Msh6* heterozygous mutants**

Figure 7-20 *Msh2* and *Msh6* heterozygous mutants

Array CGH analysis of nine mutant clones which show a heterozygous deletion covering *Msh2* and *Msh6* genes on chromosome 17. Chromosomes are also placed horizontally in ascending order followed by chromosomes X and Y. The Y axis is the Log_2(Mutant DNA signal/Reference DNA signal) of the BAC probes. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of the BAC clone probes. The red line indicates the Log_2(Reference DNA signal/Mutant DNA signal) of the BAC clone probes from the reciprocal hybridization experiment. The Log_2(ratio) indicates an amplification when it is above +0.29 or a deletion when it is below -0.29. Besides the heterozygous deletions covering *Msh2* and *Msh6*, there are other genomic DNA copy number changes. For example, D6 appears to have a monosomy on chromosomes 2, 9, 13 and 16. Some small homozygous deletions (single BAC) deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are associated with chromosome Y loss. This further confirms the assumption that these BACs should be annotated to chromosome Y. The array profiles of clone F9 and F16 are highly similar. As a clonal relationship has been established by Southern blot between these two mutants, this is an expected result. This result further confirms the clonal relationships of other mutants built by Southern blot and also verifies the reliability of the array CGH analysis.

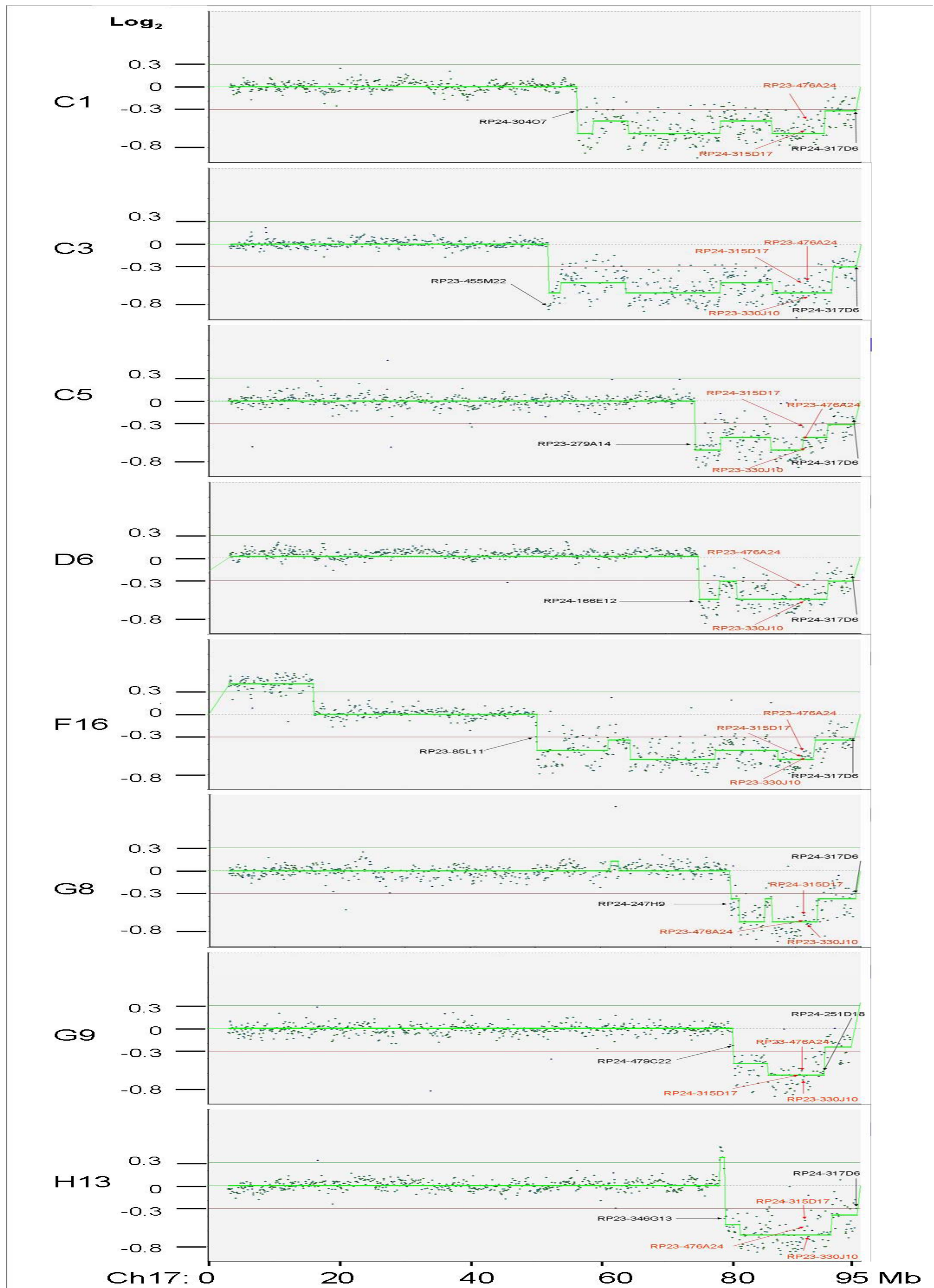

Figure 7-21 ***Msh2*** **and** ***Msh6*** **heterozygous mutants – chromosome 17**

Figure 7-21 *Msh2* and *Msh6* heterozygous mutants – chromosome 17

The array CGH profiles on chromosome 17 of eight clones (F9 and F16 is clonal thus F9 is not shown) with heterozygous deletions covering the *Msh2* and *Msh6* genes are shown. A light green line indicates the average Log_2(Mutant DNA signal/Reference DNA signal) of regions. BAC RP24-315D17 (red font) covers the *Msh2* gene, while BACs RP23-476A24 and RP23-330J10 (red font) represent a region covering the *Msh6* gene. Log_2(ratio) of these three BACs is above -1 and below zero, therefore the two genes are regarded as heterozygous. The starting BAC probe and the ending BAC probe of each deleted region are shown in a black font. BAC RP24-317D6 is the most distal BAC of chromosome 17 on the 200 kb resolution array CGH.

Table 7-5 Summary of heterozygous deletions covering *Msh2* and *Msh6*

Clone	Deleted region starting BAC (Mb)	Deleted region ending BAC (Mb)	Deletion length (Mb)	Average Log_2(ratio) over deleted region
C1	RP24-304O7 (55.4)	RP24-317D6 (95.1)	39.7	-0.51
C3	RP23-455M22 (51.3)	RP24-317D6 (95.1)	43.8	-0.59
C5	RP23-279A14 (72.2)	RP24-317D6 (95.1)	22.9	-0.52
D6	RP24-166E12 (72.9)	RP24-317D6 (95.1)	22.2	-0.43
F9	RP23-85L11 (49.1)	RP24-317D6 (95.1)	46.0	-0.51
F16	RP23-85L11 (49.1)	RP24-317D6 (95.1)	46.0	-0.47
G8	RP24-247H9 (77.3)	RP24-317D6 (95.1)	17.8	-0.49
G9	RP24-479C22 (78.2)	RP24-251D18 (91.5)	13.3	-0.55
H13	RP23-346G13 (76.3)	RP24-317D6 (95.1)	18.8	-0.52

The starting and ending BACs deleted on chromosome 17 including the *Msh2* and *Msh6* genes in nine mutant clones. The rough lengths of the deletions including these two genes are between 13.3 Mb and 46 Mb. The average deletion length is 28 Mb. The average Log_2(mutant DNA/*Blm*-deficient cell DNA) of the deleted regions is between -0.59 and -0.43. Genomic positions refer to Ensembl release 48 (NCBI m37).

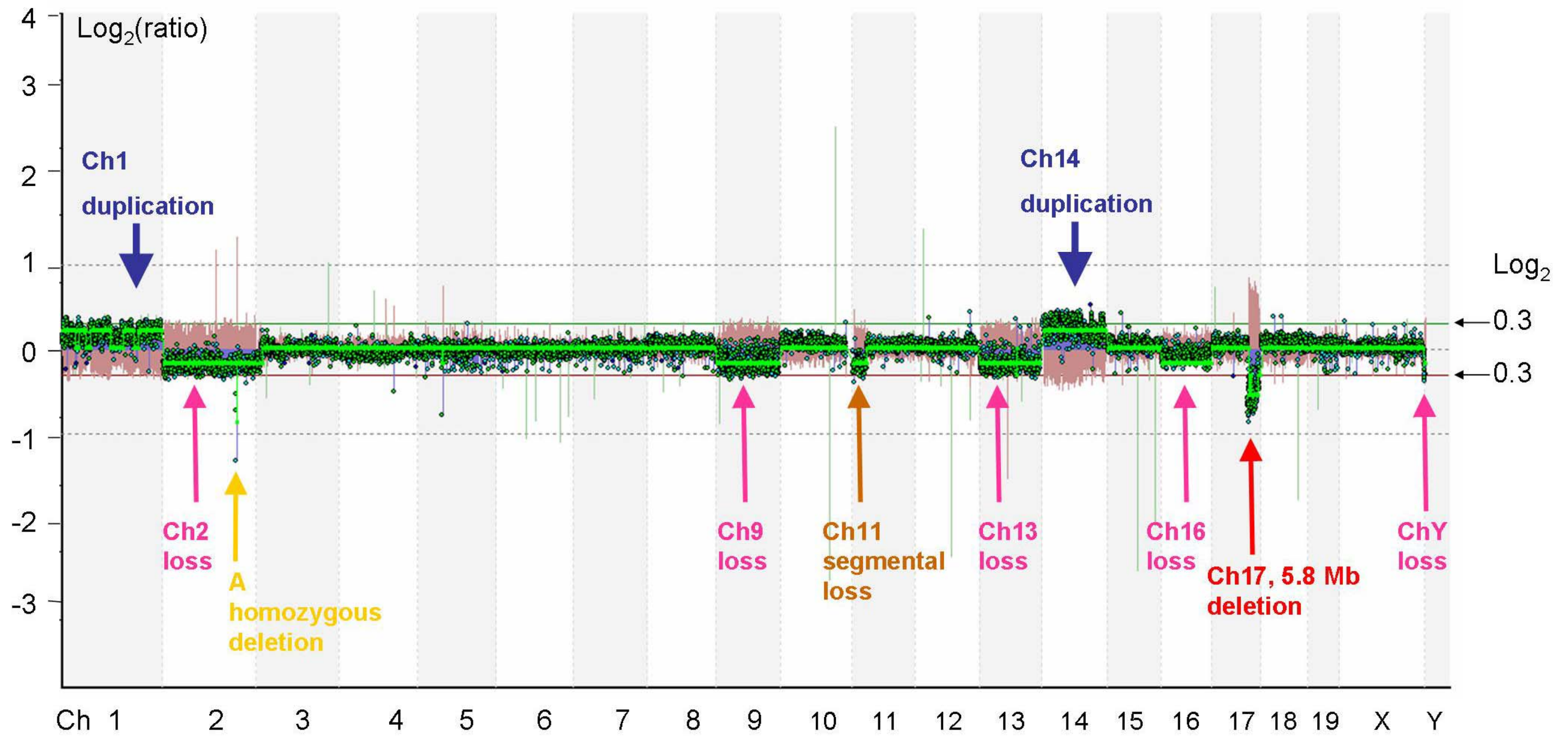

Figure 7-22 **Array CGH profile of mutant D6**

Figure 7-22 Array CGH profile of mutant D6

BACs were aligned on the X axis in ascending order according to their genomic position in every chromosome. Chromosomes were also placed in ascending order followed by X and Y. A dot indicates the middle point position of a given BAC. The Y axis lists the value of the Log_2(Mutant DNA signal/Reference DNA signal) of a given BAC probe. The Log_2(ratio) indicates an amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log_2(Reference DNA signal/Mutant DNA signal) of the BAC probes from the reciprocal hybridization experiment. A red arrow indicates a 5.8 Mb heterozygous deletion including *Msh2* and *Msh6* genes on chromosome 17. There seem to be monosomies on chromosomes 2, 9, 13 and 16 (pink arrow). The Log_2(ratio) of chromosome Y is near -0.3, indicating this clone might be a mosaic population (pink arrow). A deletion on chromosome 2 appears to be homozygous (yellow arrow). Chromosome 11 appears to contain a segmental deletion at the centromere end (brown arrow). There seem to be trisomies on chromosome 1 and 14 (blue arrow). Some duplications or deletions observed only in one array but not in the reciprocal array. These duplications and deletions might be experimental variants, thus are excluded from the final results. The grey horizontal dotted lines are for Log_2(ratio) references of either +1 or -1.

7.2.6.3 *Msh6* mutant G10

The mutant clone G10, strongly resistant to 6TG, has a 3.2 Mb long homozygous deletion on chromosome 17 (Figure 7-24). In this clone, the *Msh2* gene is not deleted because the Log_2(ratio) of its representing BAC, RP24-315D17, is zero. The Log_2(ratio) of the two BACs representing *Msh6* (Figure 7-23) is different. BAC RP23-476A24 covers the whole *Msh6* locus. It may be partially deleted because its Log_2(ratio) is close to -0.3 (Figure 7-24). While BAC RP23-330J10 covers part of the *Msh6* gene and this BAC is homozygously deleted, with the Log_2(ratio) of ~-2. Therefore the *Msh6* gene is likely to be homozygously deleted. More experiments, such as array CGH analysis with higher resolution, FISH (fluorescent in situ hybridization) or expression array analysis may be of help to clarify this. Unfortunately, this clone was contaminated thus was not able to be further analysed.

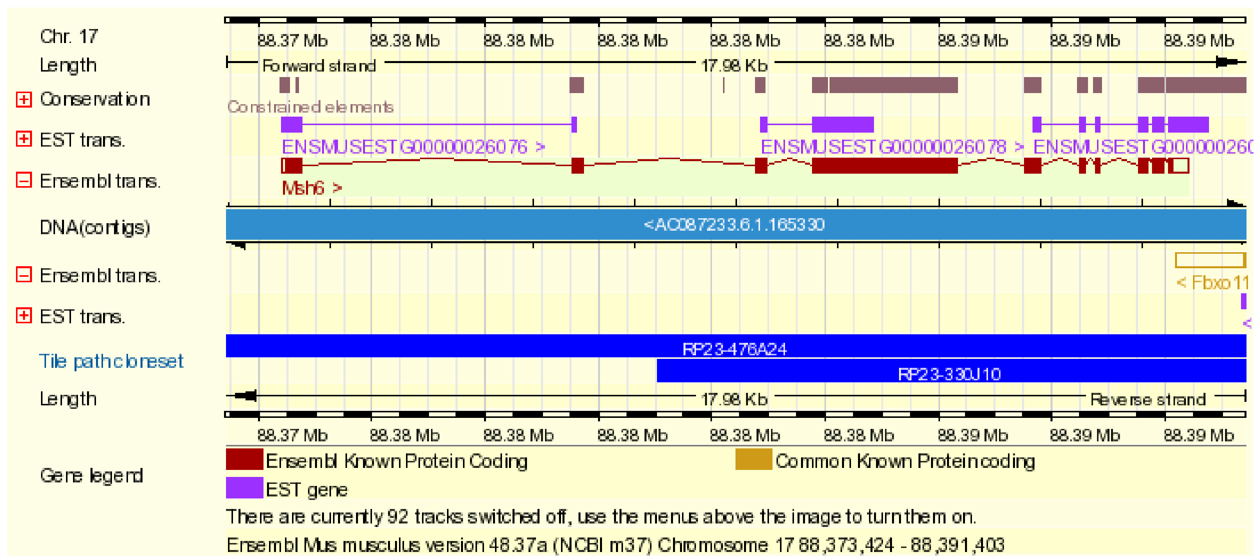

Figure 7-23 Tile path BACs at *Msh6* locus

A snapshot of the *Msh6* locus on mouse chromosome 17 (NCBI m37, Ensembl release 48). The *Msh6* transcript is shown in red boxes connected by red lines. Blue boxes indicate BACs used on the tile path array CGH (200 kb resolution). BAC RP23-476A24 covers the whole *Msh6* gene while RP23-330J10 covers part of the *Msh6* gene.

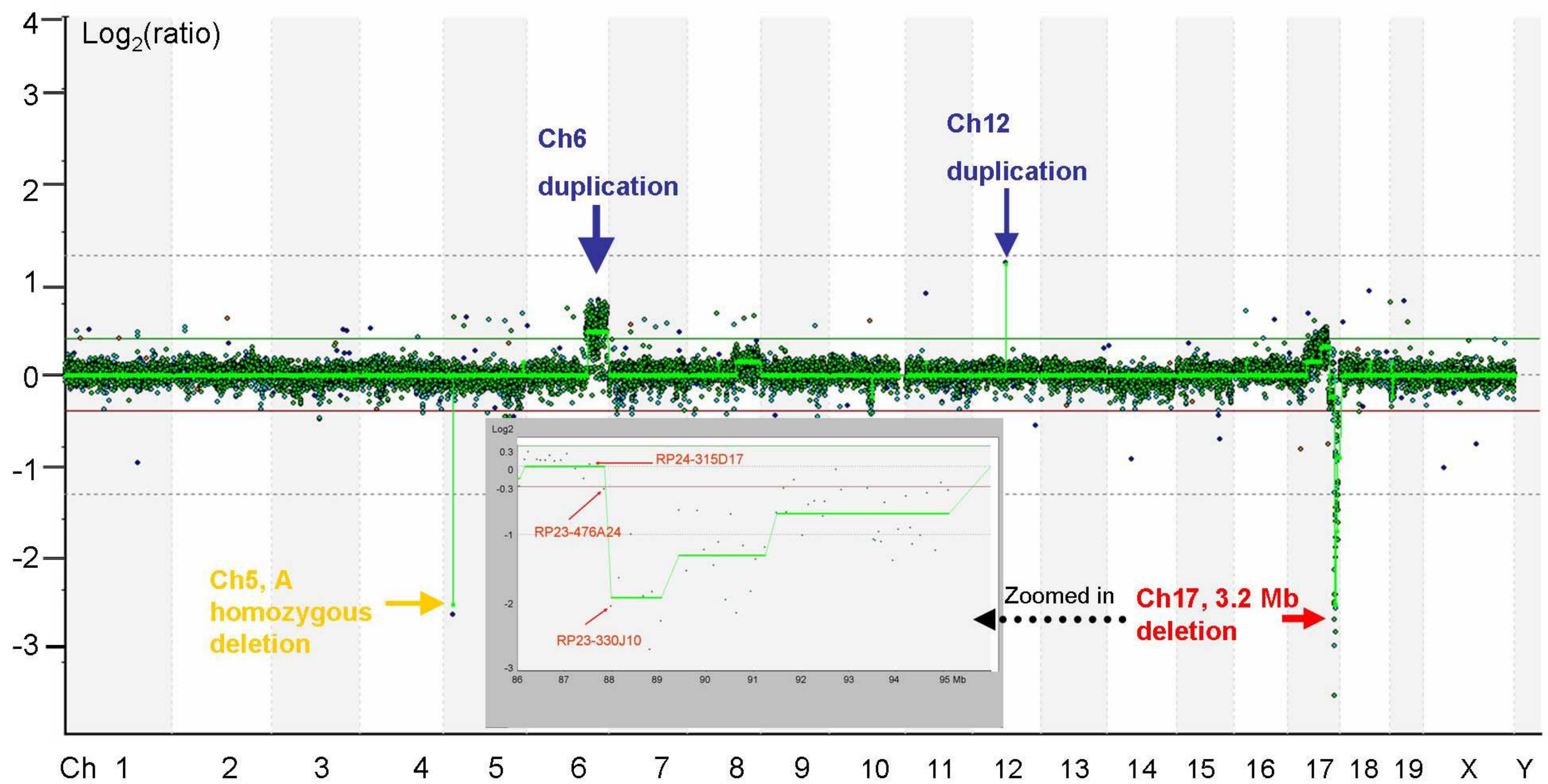

Figure 7-24 Array CGH profile of mutant G10

Figure 7-24 Array CGH profile of mutant G10

The array CGH profile of the mutant clone G10. A light green line indicates the average Log_2(Mutant DNA signal/Reference DNA signal) of regions. A 3.2 Mb long homozygous deletion is on chromosome 17 (red arrow). A zoomed-in figure of this deleted region is shown. BAC RP24-315D17 (red font) covers the *Msh2* gene and its Log_2(ratio) is zero, indicating the copy number of this BAC remains unchanged. BACs RP23-476A24 (red font) covers the whole *Msh6* gene region (Figure 7-23) while its Log_2(ratio) is about -0.3. RP23-330J10 (red font) represent part of the *Msh6* gene and its Log_2(ratio) is close to -2. In addition to this homozygous deletion, there is a homozygous deletion on chromosome 5 (single BAC, yellow arrow) and two duplications (blue arrows) on chromosome 6 (Mb) and 12 (single BAC). However, change of one BAC needs to be further examined because it can also be experimental variations.

7.2.6.4 A common deletion on chromosome 14

It was believed that 6TG resistance associated genes can be mutated for multiple times in the mutation pools and can be isolated. The homozygous and heterozygous mutations in the *Msh2* and *Msh6* genes described previously confirmed this assumption. Thus, frequently appeared mutations need to be investigated. It was found that eight clones, including B2, E1, E2, E4, E5, F2, G4 and H6 contain a common heterozygously deleted region on chromosome 14 (Figure 7-25). Five of them (E1, E4, E5, F2 and H6) are strongly resistant to 6TG, while the other three (B2, E2 and G4) are weakly resistant to 6TG. This common deleted region is between 107.9 Mb (RP24-115C5) and 119.6 Mb (RP23-407B7) on chromosome 14, indicated by the starting and the ending BAC probes.

The genes within this region were extracted from the Ensembl database (Table 7-6). Some clues were found to establish links between genes in this list and DNA repair function. The gene of ATP-binding cassette, sub-family C (CFTR/MRP), member 4 (*Abcc4*), the mouse homologue of human multidrug resistance-associated protein 4 gene (*MRP4*), is in this region. Mice deficient for this gene are more vulnerable to the nucleotide analogue, 9'-(2'-phosphonylmethoxyethyl)-adenine (PMEA), suggesting the protective function of ABCC4 in defence of DNA damage (Belinsky *et al.* 2007). Therefore, the *Abcc4* gene is worth further investigating.

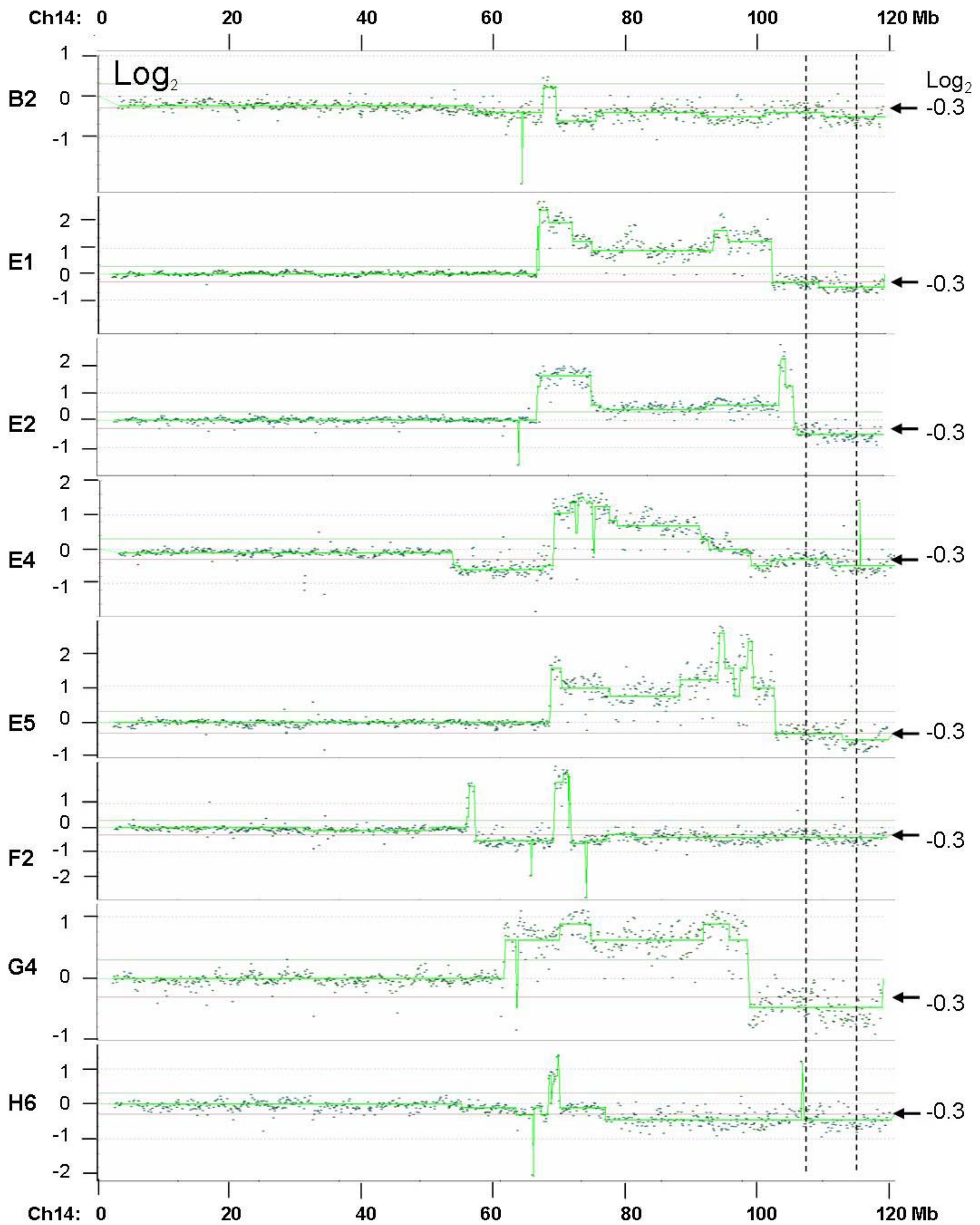

Figure 7-25 A common heterozygous deletion on chromosome 14

The array CGH profiles on chromosome 14 of the mutant clones B2, E1, E2, E4, E5, F2, G4 and H6. Two dotted lines indicate the common heterozygous deletion in the eight clones. The Log_2(ratio) of these deletions is below -0.3 and above -1.

Table 7-6 Genes in the common heterozygous deletion on chromosome 14

MGI symbol	Description
Slitrk1	SLIT and NTRK-like family, member 1 [MGI:2679446]
Slitrk6	SLIT and NTRK-like family, member 6 [MGI:2443198]
Slitrk5	SLIT and NTRK-like family, member 5 [MGI:2679448]
Tpm3	tropomyosin 3, gamma [MGI:1890149]
Q78E13_MOUSE	RNA for type IIB intracisternal A-particle (IAP) element encoding integrase, clone 111. (Fragment). Source: Uniprot/SPTREMBL Q78E13
Gpc5	glypican 5 [MGI:1194894]
Gpc6	glypican 6 [MGI:1346322]
Dct	dopachrome tautomerase [MGI:102563]
Tgds	TDP-glucose 4,6-dehydratase [MGI:1923605]
Gpr180	G protein-coupled receptor 180 [MGI:1930949]
Sox21	SRY-box containing gene 21 [MGI:2654070]
Abcc4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4 [MGI:2443111]
A830021K08Rik	RIKEN cDNA A830021K08 gene [MGI:2443998]
Q8C3Q2_MOUSE	13 days embryo lung cDNA, RIKEN full-length enriched library, clone:D430043M02 product:hypothetical protein, full insert sequence. Source: Uniprot/SPTREMBL Q8C3Q2
Cldn10	claudin 10 [MGI:1913101]
Dnajc3	DnaJ (Hsp40) homolog, subfamily C, member 3 [MGI:107373]
Dzip1	DAZ interacting protein 1 [MGI:1914311]
Ugcgl2	UDP-glucose ceramide glucosyltransferase-like 2 [MGI:1913685]
Hs6st3	heparan sulfate 6-O-sulfotransferase 3 [MGI:1354960]

These genes locate between 107.9 and 119.6 Mb on chromosome 14 and are sorted in the direction from centromere to telomere.

7.2.6.5 The Other mutant clones

Of the thirty-four mutant clones analysed using array CGH, twenty-four have been described previously (summarized in Table 7-7). Although homozygous deletion covering the MMR genes of *Msh2* and *Msh6* were previously identified, other MMR genes, such as *Mlh1* and *Pms2*, failed to be identified in the other ten 6TG-isolated mutant clones (B3, B7, E3, E9, F1, F6, G6, H1, H3 and H5). One possible reason is that the adjacent genes of *Mlh1* and *Pms2* are essential to ES cell viability. Loss of function of these essential genes can result in cell lethality. Based on the data described previously, the homozygous deletions can be three to four million base pairs in size; and the adjacent regions of *Mlh1* and *Pms2* genes are gene-dense, therefore the probability to delete adjacent essential genes together with the MMR genes is high. The deficiency of some genes adjacent to *Mlh1* or *Pms2*, for instance *Tdgf1* (Ding *et al.* 1998), *Ubp1* (Parekh *et al.* 2004), *Rac1* (Sugihara *et al.* 1998) and *Trrap* (Herceg *et al.* 2001) causes embryonic lethality in knockout mice. These evidence implies that the deficiency of these genes may lead to cell lethality. Moreover, *Tdgf1* is 300 kb upstream of *Mlh1* and *Ubp1* is 2.7 Mb downstream of *Mlh1*; *Rac1* is 400 kb upstream of *Pms2* and *Trrap* is 800 kb downstream of *Pms2*. Within such short distances, homozygous deletions are able to delete these vital genes together with *Mlh1* or *Pms2,* thus *Mlh1* or *Pms2* deficient mutants can not be isolated.

The array CGH profiles of the other ten mutant clones are shown in Figure 7-26–35. Among them, the mutant clone B3 is weakly resistant to 6TG while the other nine clones are strongly resistant to 6TG (Table 7-2, C). The loss of chromosome Y was observed in seven out of ten mutants (B7, E3, E9, F1, F6, H1 and H3). Some single BAC homozygous-like deletions were seen on chromosome 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X. These deletions were found to be associated with chromosome Y loss. Thus, they were assumed to belong to chromosome Y but were incorrectly annotated to other chromosomes. Besides these single BAC deletions, some other single BAC homozygous deletions were observed, for example those on chromosomes 3, 15 and 17 of the clone G6, and that on chromosome 6 of the clone H5. However, single BAC changes are not convincing data. One needs to be cautious with them. Due to the complexity of the duplications and deletions in these clones, it is difficult to conclude which gene is responsible for the 6TG resistance. In an attempt to answer this question with information at transcriptional level, expression array analysis using an Illumina® platform was conducted (Chapter 8).

Table 7-7 Summary of the mutants analysed using array CGH

Types of mutation	ID of mutant clones	Number of mutant clones
Homozygous deletion of *Msh2* and *Msh6*	B6, D4, D8, F4, H14	5
Homozygous deletion of *Msh6*	D1	1
Heterozygous deletion of *Msh2* and *Msh6*	C1, C3, C5, D6, F9, F16, G8, G9, H13	9
A potential homozygous deletion of *Msh6*	G10	1
A common heterozygous deletion on ch14	B2, E1, E2, E4, E5, F2, G4, H6	8
Others	B3, B7, E3, E9, F1, F6, G6, H1, H3, H5	10
Total number		34

Types of the thirty-four mutants analysed using array CGH.

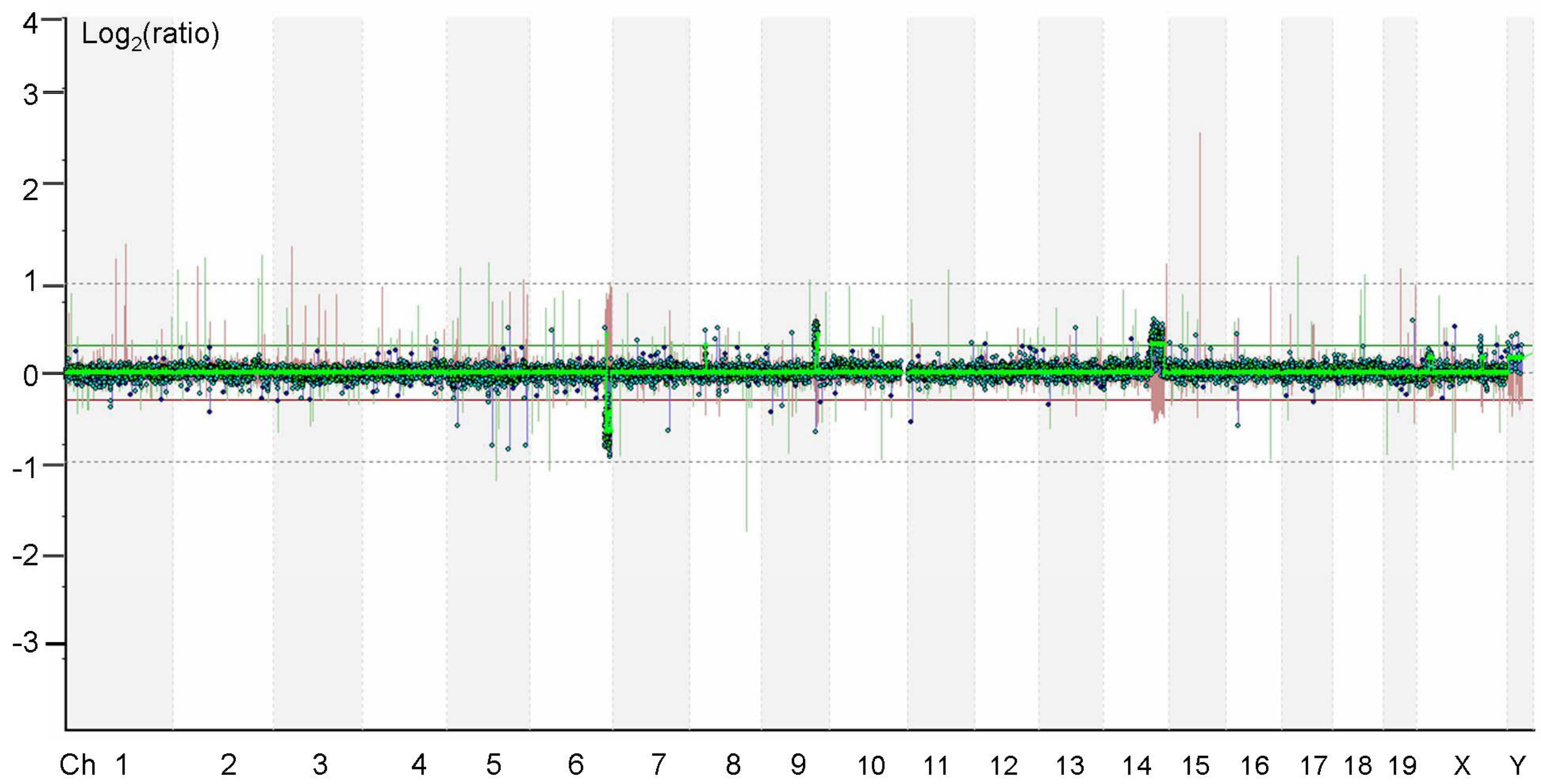

Figure 7-26 **Array CGH profile of mutant B3**

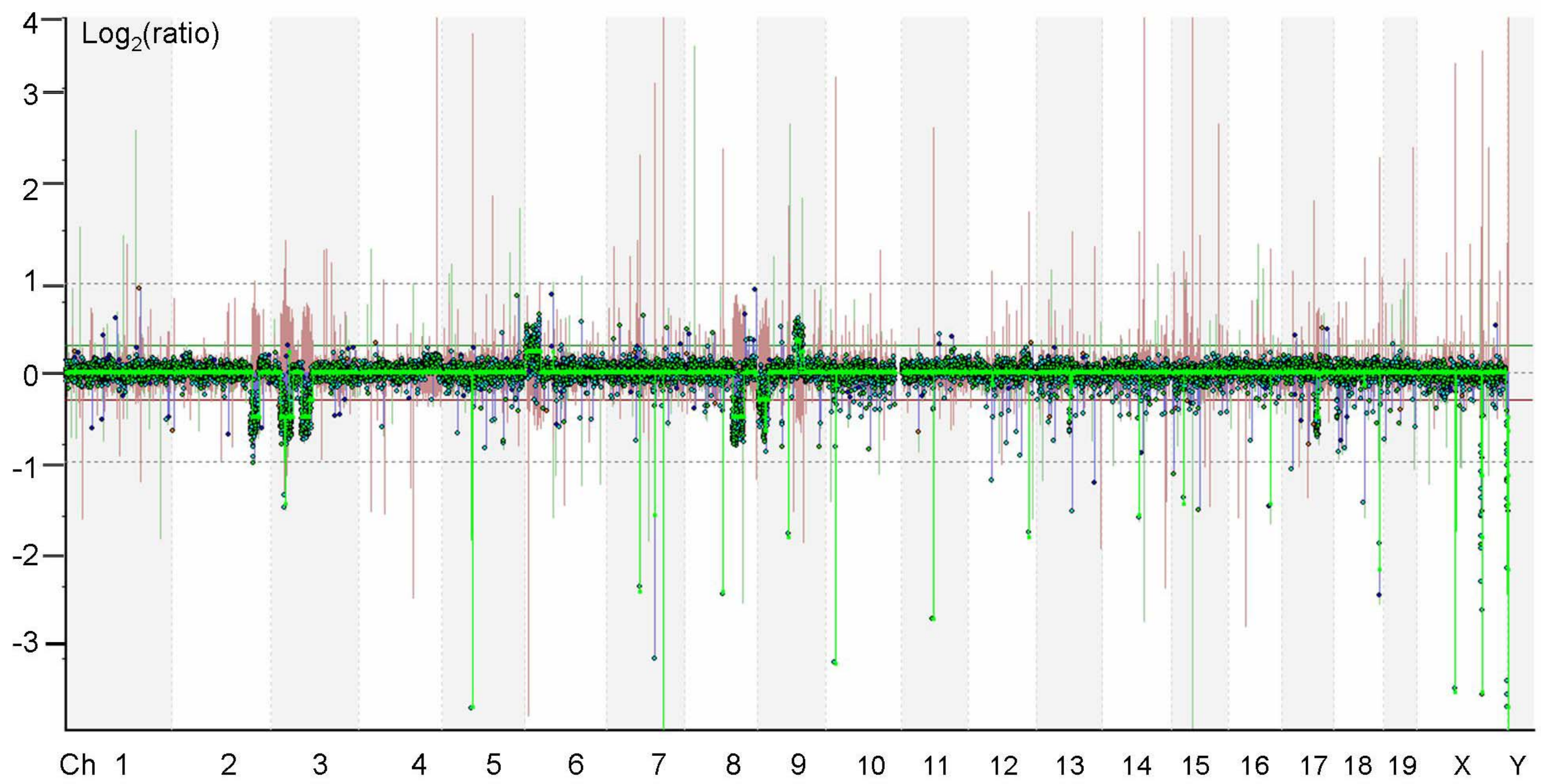

Figure 7-27 Array CGH profile of mutant B7

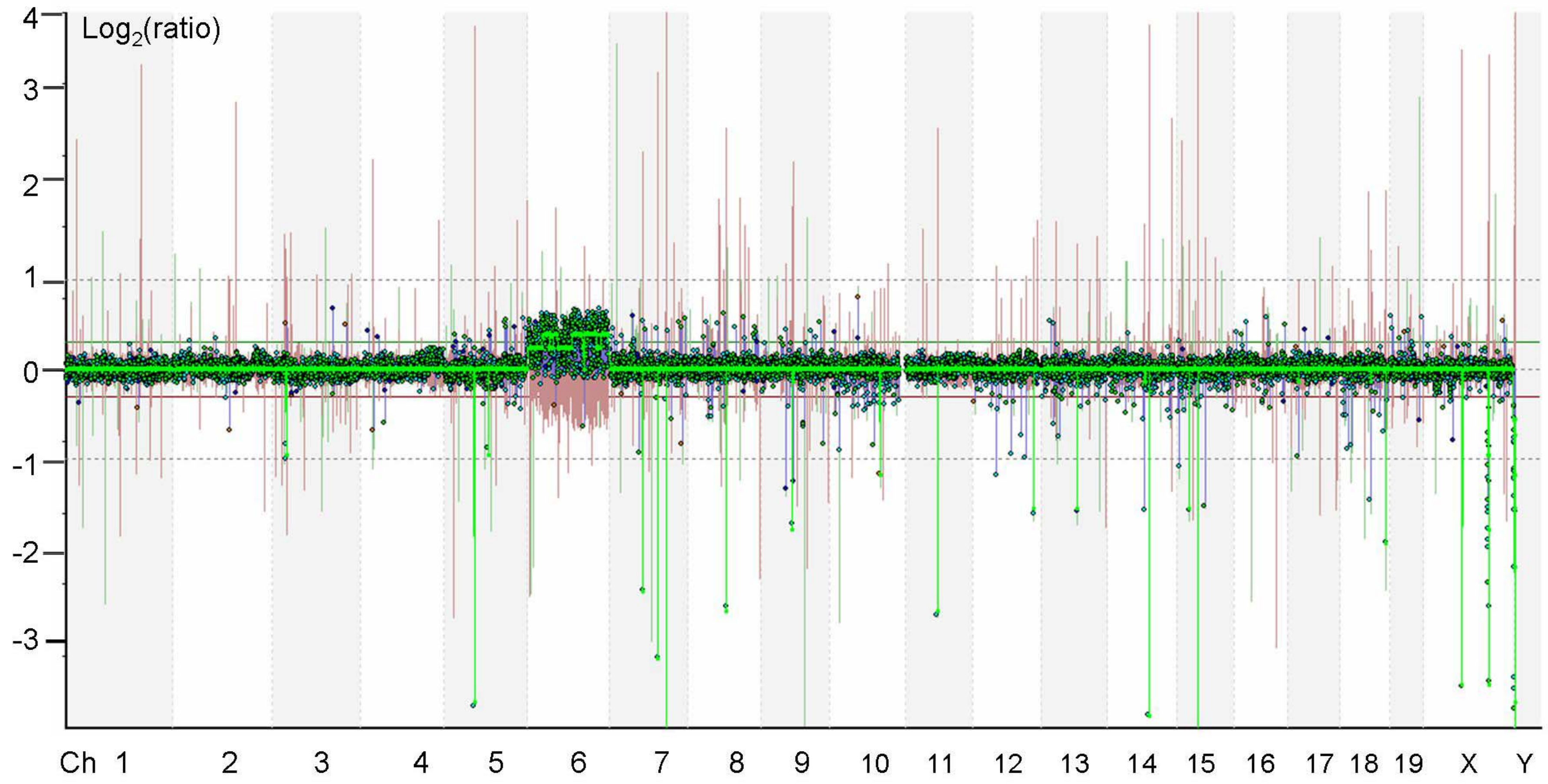

Figure 7-28 **Array CGH profile of mutant E3**

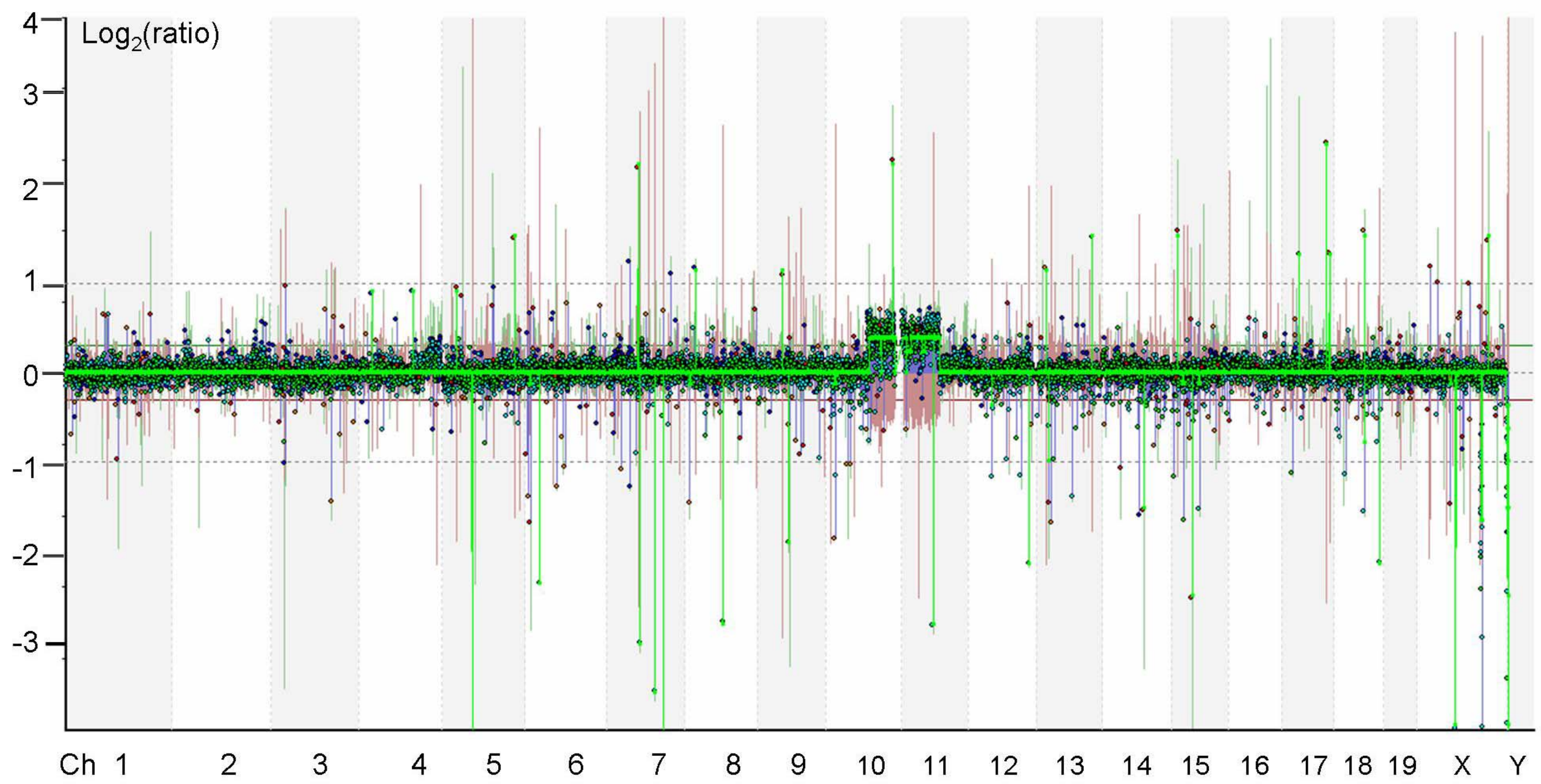

Figure 7-29 Array CGH profile of mutant E9

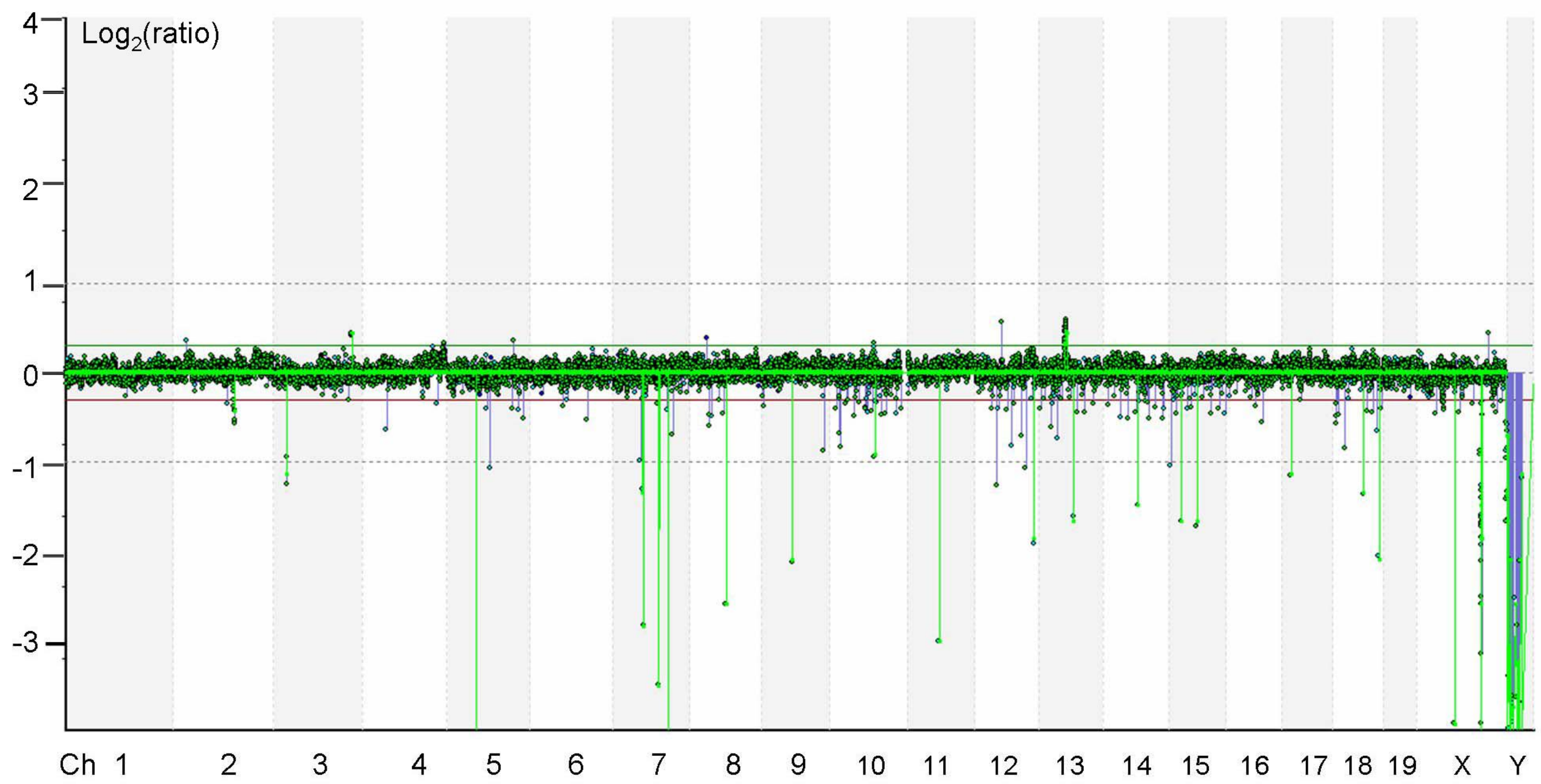

Figure 7-30 **Array CGH profile of mutant F1**

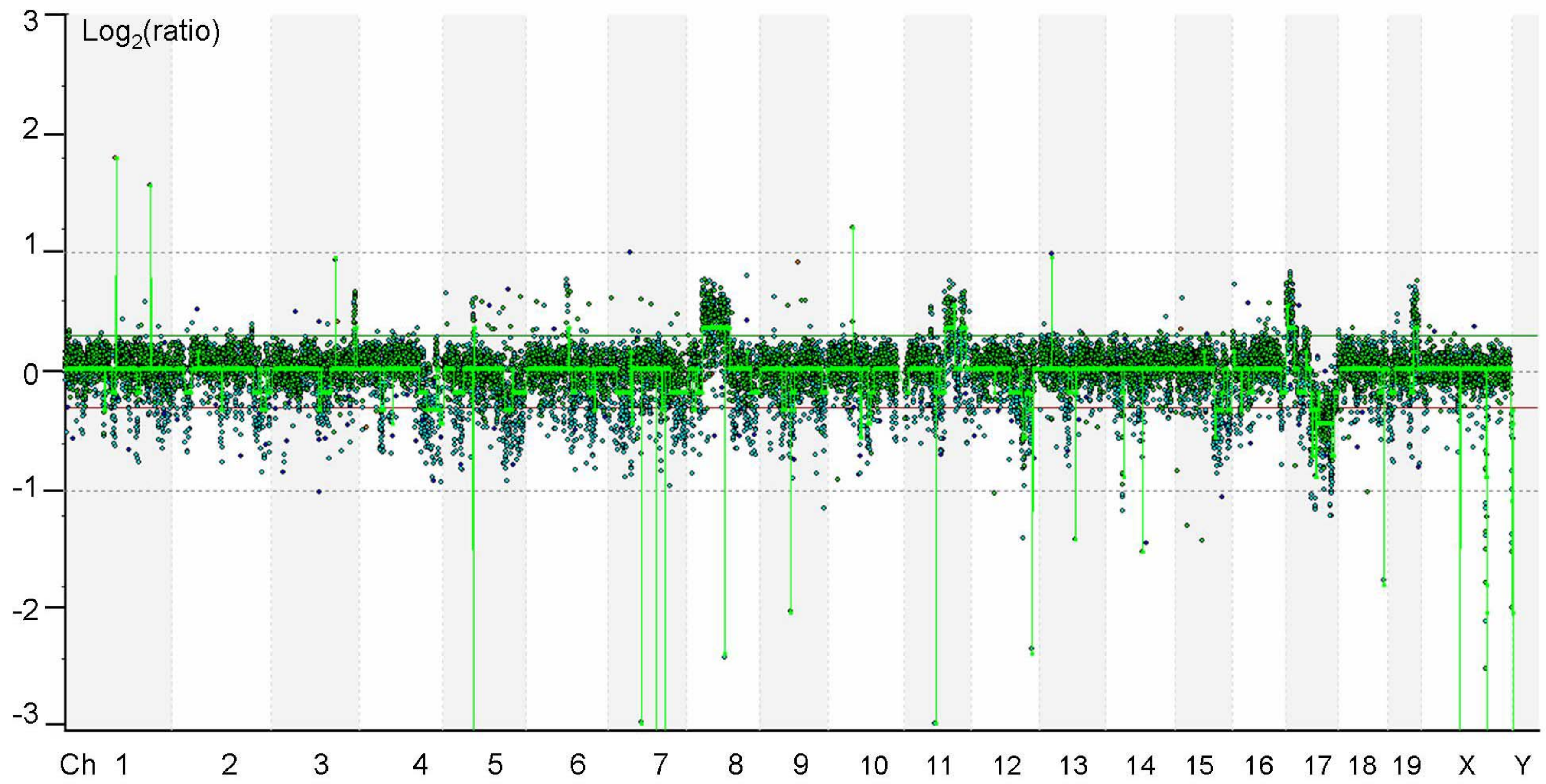

Figure 7-31 **Array CGH profile of mutant F6**

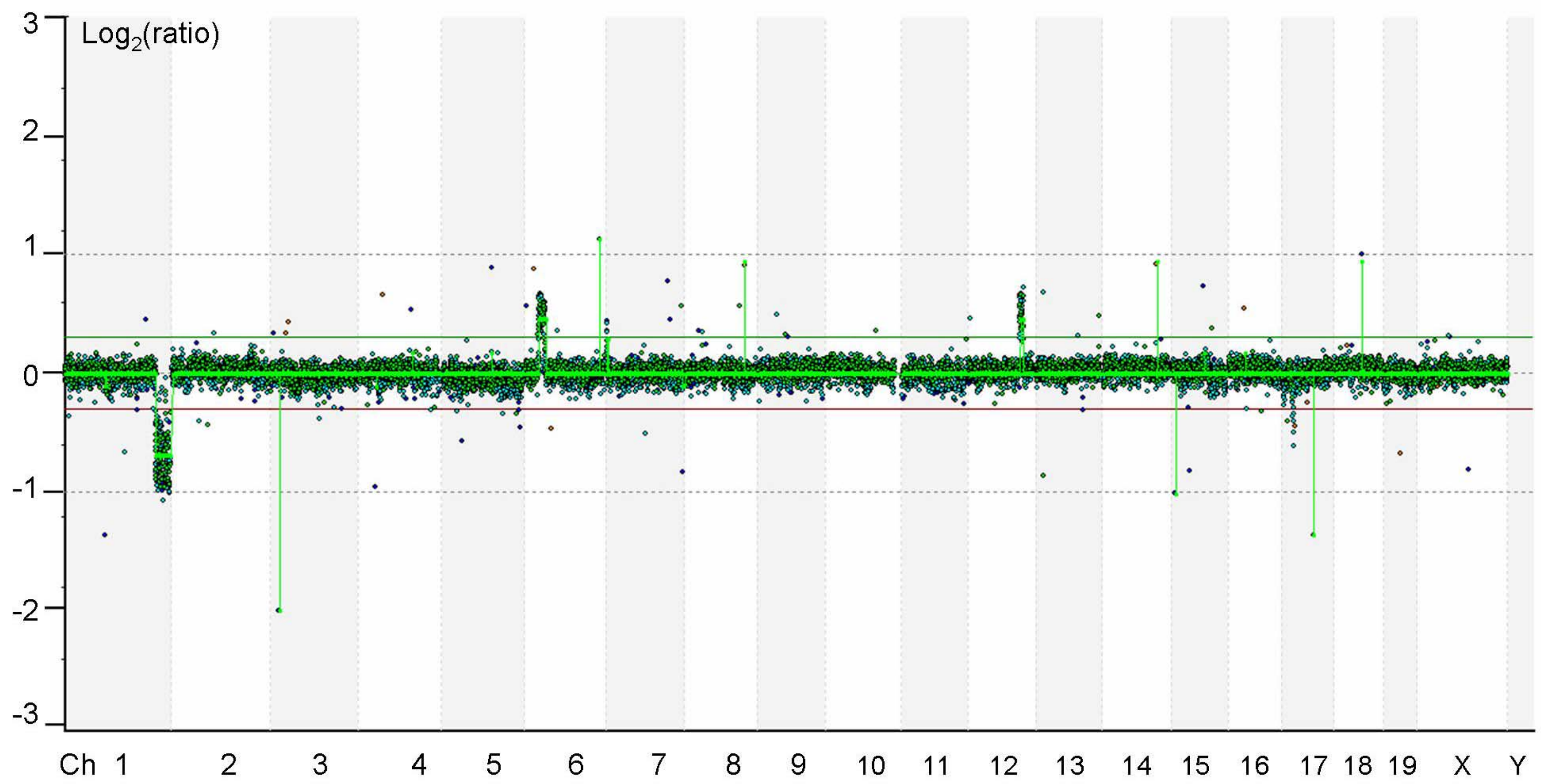

Figure 7-32 **Array CGH profile of mutant G6**

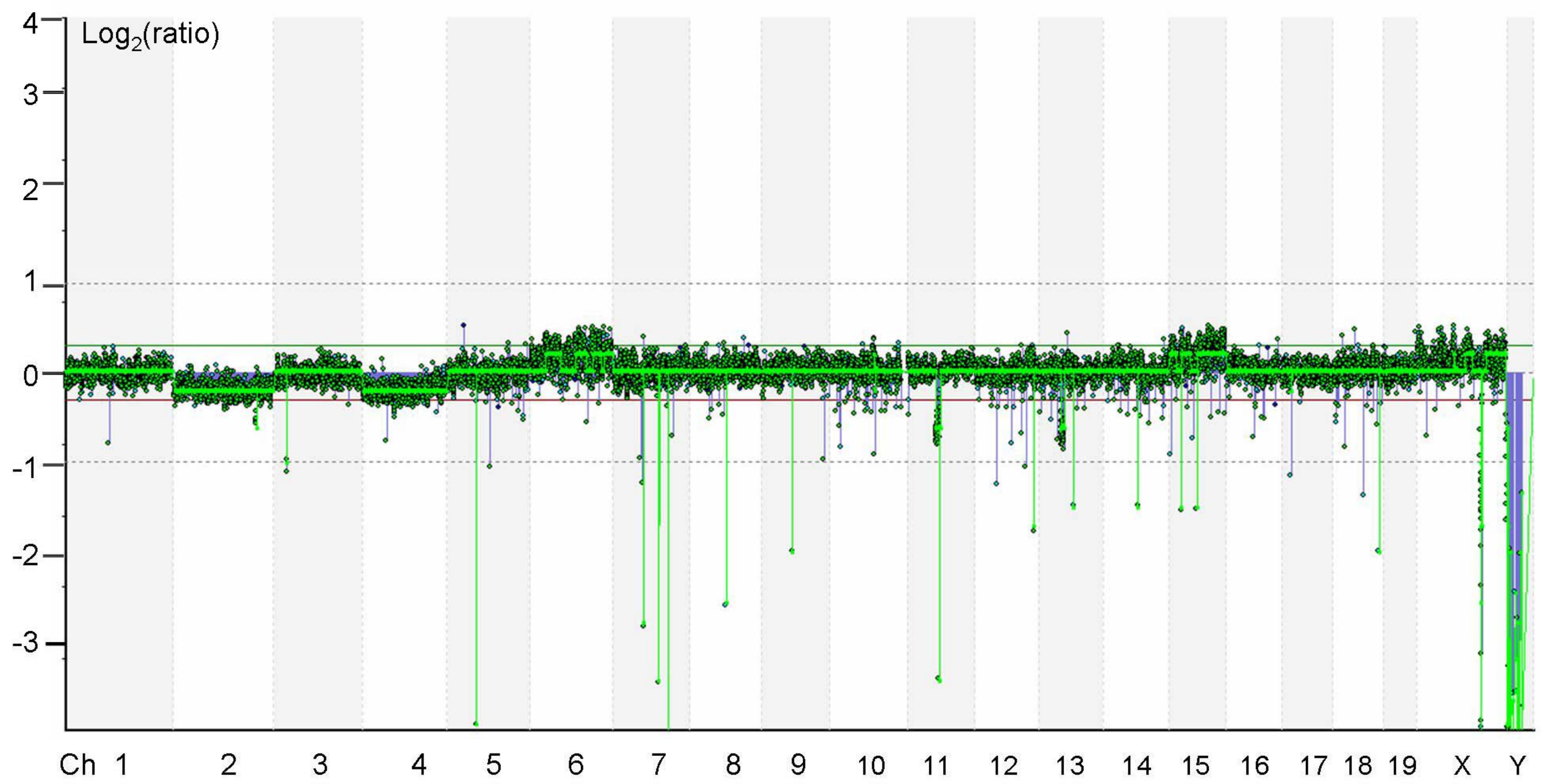

Figure 7-33 **Array CGH profile of mutant H1**

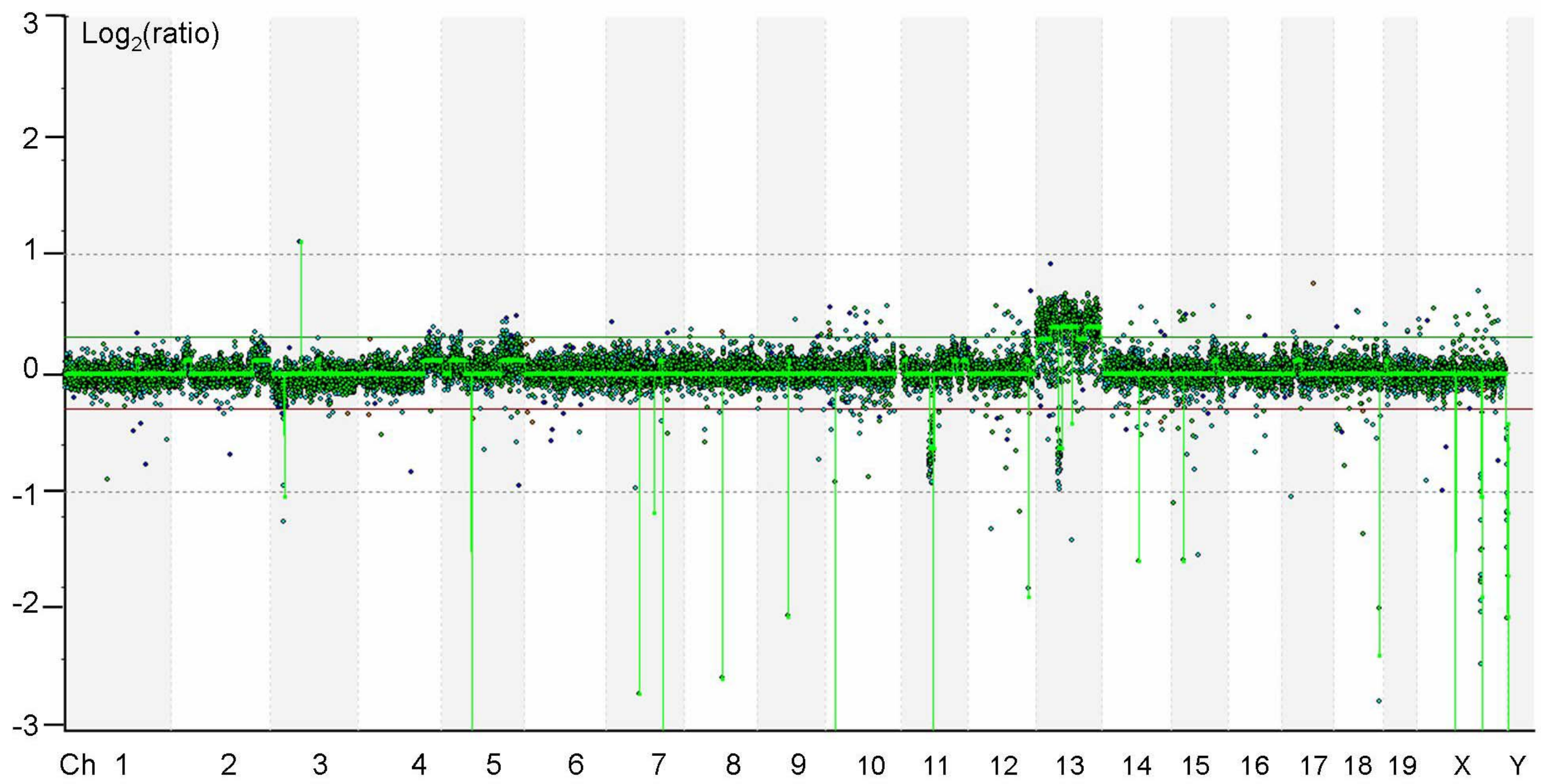

Figure 7-34 Array CGH profile of mutant H3

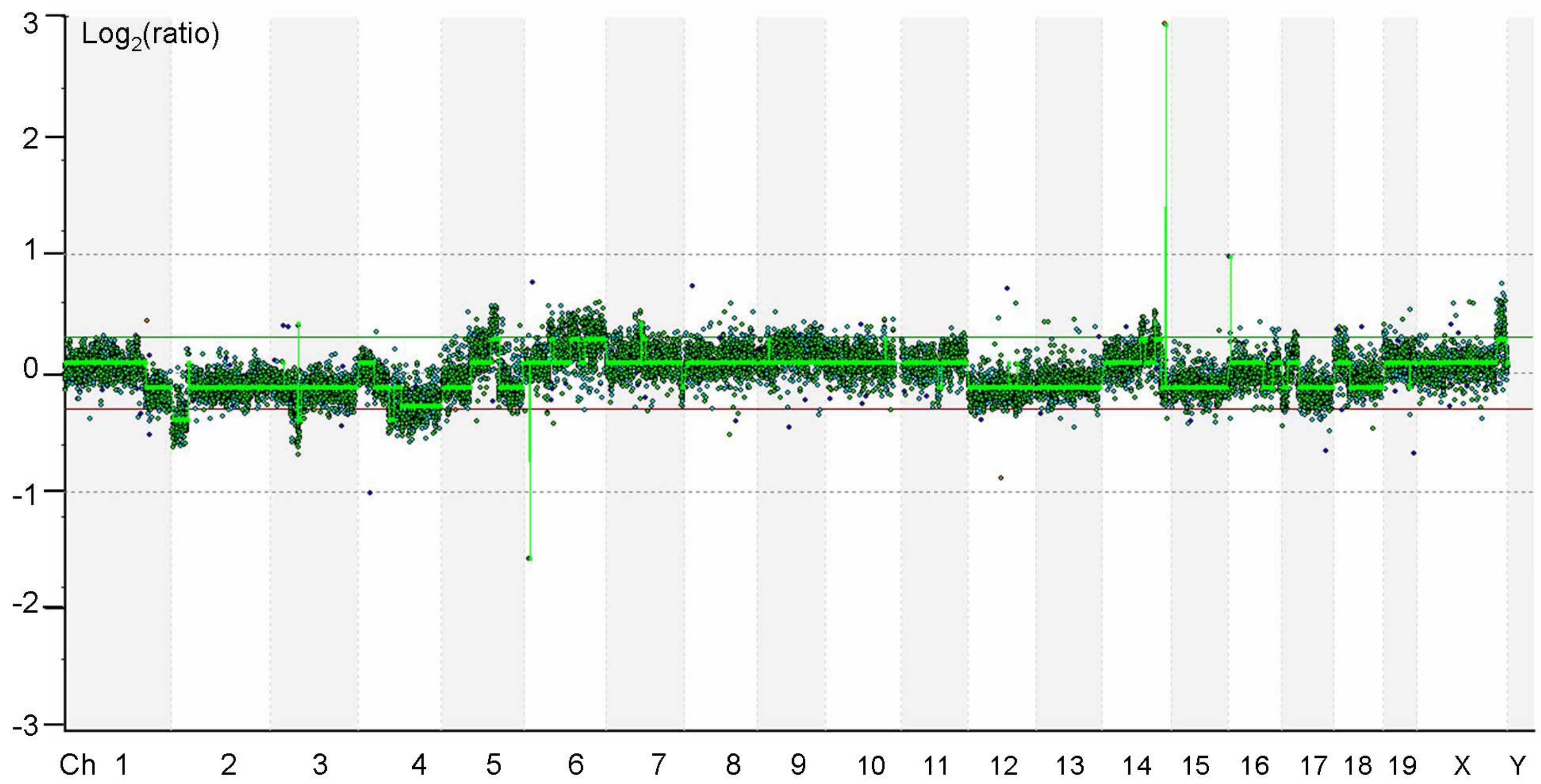

Figure 7-35 **Array CGH profile of mutant H5**

7.3 Conclusion and discussion

7.3.1 Advantages of establishing clonal relationships

Mutations generated by insertional mutagenesis methods are uniquely tagged and can be identified by molecular methods. However, mutations generated by irradiation cannot be so easily identified. The analysis methods for irradiation-induced mutations are genome-wide array CGH and expression arrays. These approaches are expensive. To prevent redundant experiments, molecular tags were introduced to distinguish clonal relationships between mutants and identify independent clones. A retroviral vector was incorporated into the cells before they were irradiated. By controlling the infection efficiency, it was simple to ensure each cell had a unique tag. In order to guarantee each cell contains one tag after irradiation, the complexity of each mutation pool was limited to approximate one thousand cells per pool, one sixth of the total number of tags before irradiation. Among 49 mutants with strong 6TG resistance, twenty-seven were identified as unique clones after analysing clonal relationships. Thus, only 27 unique clones needed to be analysed on arrays.

7.3.2 Mutations were identified in *Msh2* and *Msh6* MMR genes

Within 27 unique 6TG fully resistant mutants (Table 7-3), 22% (6/27) had homozygous deletions at MMR gene loci, including *Msh2* and *Msh6*. These six mutants, B6, D1, D4, D8, F4, and H14, were from four mutation pools. These *Msh2* and *Msh6* deletions ranged from 2.6 to 4.4 Mb. These samples provide size information about homozygous deletions generated by irradiation and the LOH events of the *Blm*-deficient cells. The average size of these six homozygous deletions is 3.6 Mb. It is adequate for 200 kb resolution array CGH to detect such deletions.

The mutant of other MMR genes, for instance *Mlh1* or *Pms2,* was not isolated. This may be due to presence of genes vital to cell viability around these two genes. *Tdgf1* is 300 kb upstream of *Mlh1* and *Ubp1* is 2.7 Mb downstream of *Mlh1*; *Rac1* is 400 kb upstream of *Pms2* and *Trrap* is 800 kb downstream of *Pms2*. The deficiency of these four genes causes embryonic lethality in knockout mice (Ding *et al.* 1998; Herceg *et al.* 2001; Parekh *et al.* 2004; Sugihara *et al.* 1998), implying that the deficiency of these genes may lead to cell lethality. It is known that homozygous deletions can be three to four million base pairs in size (Table 7-4), thus homozygous deletions are able to delete these vital genes together with *Mlh1* or *Pms2.*

In addition to the six clones with homozygous deletions, nine clones (C1, C3, C5, D6, F9,

F16, G8, G9 and H13) were identified which had heterozygous deletions of the same region on chromosome 17. The Log_2(ratio) of these deletions is between -0.5 and -0.7 (Figure 7-20). These nine deletions range from 13.3 Mb to 46 Mb in size (Table 7-5), much larger than the sizes of homozygous deletions covering the *Msh2* and *Msh6* genes (Table 7-4). The average length of heterozygous deletion is 28 Mb.

The region, including the *Msh2* and *Msh6* genes, is located at the 88 Mb position of the 95 Mb long chromosome 17 thus is near to telomere. It was believed that such a region is easy to be lost through homologous recombination. It seems all heterozygous deletions (Figure 7-21), except that in the mutant G9, in the eight clones extend to the last BAC probe on chromosome 17 (RP24-317D6), suggesting that these eight clones lost the whole distal part of chromosome 17. Although the heterozygous mutant clones still have one copy of genes *Msh2* and *Msh6*, six of them (C1, C5, F16, G8, G9 and H13) have shown strong 6TG resistance. It is assumed that the undeleted allele of either *Msh2* or *Msh6* has a small mutation, which results in a loss-of-function of either *Msh2* or *Msh6.* It is anticipated that expression array analysis will confirm that the transcripts of *Msh2* and *Msh6* are affected. Loss of distal part of chromosome 17 was also identified in the six mutants with homozygous *Msh2* and *Msh6* deletions (Figure 7-18).

Monosomies and trisomies were found in the heterozygous *Msh2* and *Msh6* mutants. For example, clone D6 appears to have monosomies of chromosomes 2, 9, 13 and 16 and trisomies on chromosome 1 and 14 (Figure 7-22). Chromosome 11 appears to contain a segmental deletion at the centromere end. However, the Log_2(ratio) of these changes is nearer to zero than that of other heterozygous deletions and duplications, indicating that this clone might be a mosaic population or a tetraploid cell or both. There are some evidence to support this hypothesis: firstly, the Log_2(ratio) of chromosome Y is near -0.3, higher than that of the normal heterozygous deletions. The ES cells are XY cells. If chromosome Y is lost, the Log_2(ratio) of chromosome Y on the array CGH should be like those of homozygous deletions. If one chromosome Y remains in a tetraploid cell, the Log_2(ratio) of it should be close to that of heterozygous deletions. However, the Log_2(ratio) of chromosome Y in clone D6 is near 0.3, suggesting that D6 is an impure clone. Secondly, monosomies are very rare in ES cells. Thus, the clone might be a mosaic population of normal ploidy cell with tetraploid cells with one copy of chromosome loss (Log_2(¾)= -0.42). Thirdly, the Log_2(ratio) of chromosome 1 and 14 is less than 0.32, the Log_2(ratio) of a pentaploid chromosome in a tetraploid cell (Log_2(5/4)), indicating this clone is mosaic.

7.3.3 Clues of potential MMR genes from a common deletion

The six homozygous and nine heterozygous mutations in the *Msh2* and *Msh6* genes demonstrate that the mutation of an MMR gene can be isolated in multiple clones. This kind of frequently appeared deletion needs to be investigated because 6TG resistance associated genes may locate in such regions. In the 6TG-isolated mutants, a common heterozygously deleted region on chromosome 14 was identified in eight mutant clones (B2, E1, E2, E4, E5, F2, G4 and H6) (Figure 7-25). This common deleted region is between 107.9 Mb (RP24-115C5) and 119.6 Mb (RP23-407B7) on chromosome 14. *Abcc4*, one of the genes in this region, was found to be associated with nucleotide analogue resistance, which is probably related to MMR. The *Abcc4* gene encodes ATP-binding cassette, sub-family C (CFTR/MRP), member 4 (*Abcc4*). Mice deficient for this gene exhibit increased lethality when treated with a nucleotide analogue, 9'-(2'-phosphonylmethoxyethyl)-adenine (PMEA) (Belinsky *et al.* 2007), suggesting that the *Abcc4* gene may play a role in the toxicity pathway of 6TG. Therefore, the deletion in a common region provides a clue to investigate potential MMR-related genes.