CHAPTER 8. Expression analysis of the 6TG^R mutants

8.1 Introduction

Some heterozygous deletions of the *Msh2* and *Msh6* genes are strongly resistant to 6TG. It was believed that small mutations exist in the undeleted alleles of these genes, which lead to the 6TG resistance. To test this assumption, expression array analysis was used to examine the presence of their transcripts. If the transcriptional level of heterozygously deleted genes is similar to that of homozygously deleted genes, the presence of such mutations is indirectly confirmed. Moreover, the transcriptional loss of homozygous deletions can be used to calibrate expression arrays. They can serve as negative controls of signals from non-expressed genes.

In addition, expression micro-arrays generate information not only about the presence of a given transcript, but also the expression level of a given gene. When probes on an array represent all of the genes in a genome, expression array can be used to screen transcriptional variations of the mutants. An Illumina® Mouse-6 Expression BeadChip was used to identify whole genome expression variations in the 6TG-isolated mutants.

8.2 Results

8.2.1 Expression analysis

Twenty-five unique clones (Table 7-3, without clone F16 and G10) with strong 6TG resistance were analysed not only by array CGH, but also by expression array. Homozygous deletions at known MMR genes, for instance *Msh2* and *Msh6*, were identified by array CGH analysis. As for the MMR genes, which are not homozygously deleted, absence of transcript in any of them can be expected to be identified through expression array analysis. This may be caused by small mutations (shorter than 100 kb) or nonsense-mediated mRNA decay in mutant clones.

Before initiating the expression array analysis on 6TG^R mutants, experiments were conducted to measure the sensitivity and accuracy of the arrays. A comparison between wild type cells (including both AB1 and AB2.2 cells) and *Blm*-deficient cells (cell line NGG5-3) was conducted to determine if there were any expression changes in the *Blm*-deficient cells. Individual MMR-deficient mutants were compared with the *Blm*-deficient cells to identify expression changes in the mutants. Downstream regulated gene expression

changes were identified by comparing and clustering mutants.

Biological replicates were used in the expression array analysis. These were RNA samples extracted from the same passage of a particular clone. Twenty-five 6TG-resistant mutant clones (three replicates each), six weakly resistant mutant clones (three replicates each), four wild type cell replicates (two in AB1 and two in AB2.2), three *Blm*-deficient cell replicates (NGG5-3) and three *Dnmt1* knockout cell replicates (ww56+*Hprt*) were used to conducted the analysis on an Illumina® platform. As described in the Methods, the ww56+*Hprt* cell line was derived from the AB2.2 cell line and has a targeted deletion of exons 2–4 in both alleles of *Dnmt1*. In addition, this cell line has an *Hprt* mini-gene, targeted at one allele of the *Gdf9* locus, providing the activity of hypoxanthine guanine phosphoribosyl transferase.

Each Illumina® array has approximately 30 features per probe and 19,400 genes on array. A total 7645 genes have more than one probe. A detection P-value was calculated for probes using BeadStudio (version 2.3.41) software. Probes with certain level of variations between its features are excluded from analysis. A detection P-value provides a measure of the probability of a measured signal being due to hybridization with Cy3 labelled complementary RNA rather than background non-specific binding. Each array experiment generated one signal value for a given feature of a probe. The expression fold change was calculated by comparing signals of one probe in replicates of one cell line and signals of the same probe in replicates of another cell line. The adjusted P-value was calculated per probe to indicate the confidence of expression fold change. The lower the adjusted P-value is, the more likely the expression fold change data is real. Five thousand probes with the lowest adjusted P-values in each pair of comparisons were selected for analysis. The highest P-value of five thousand probes in the arrays is between 10^{-4}–10^{-2}, thus gives a good confidence to the analysis. The other probes with higher adjusted P-values, even if they have a greater magnitude of signals, were excluded from the comparison analysis.

8.2.1.1 Controls of wild type cell lines – AB1 and AB2.2

AB1 and AB2.2 cell lines are both wild type and isolated from the same subline of 129SvEv mice (Simpson *et al.* 1997). The AB2.2 cell line was isolated from a mouse line that was still being backcrossed. In the AB2.2 cell line, the X-linked *Hprt* locus was inactivated by insertion of a myeloproliferative sarcoma virus (MPSV) (King *et al.* 1985; Kuehn *et al.* 1987; Lobel *et al.* 1985). MPSV is a variant of the Moloney strain of murine leukemia virus (MMuLV). By comparing probe signals between AB1 and AB2.2 cells, I could expect to

detect expression variations of both the *Hprt* gene and the MPSV or MMuLV associated elements.

Five thousand probes with the lowest adjusted P-values in the comparison were selected for analysis. The majority of genes are expressing at the same level between the AB1 and AB2.2 cell lines. There was no probe with a high fold change and high adjusted P-value. Thirteen probes (representing eleven genes) have twice the expression in AB1 than in AB2.2 (Table 8-1). Two out of the three *Hprt1* gene probes were among them, with 2.75 and 2.73 times the expression level in AB1 than in AB2.2 (Figure 8-2). Another *Hprt1* gene probe, scl0002995.1_2-S, has 1.98 times the expression level in AB1 than in AB2.2, close to the threshold of 2 times. This result is consistent with the fact that the *Hprt1* gene was inactivated in AB2.2 cells (Kuehn *et al.* 1987). Six probes were found to express at least 2 times more in AB2.2 than in AB1. These probes detect gene expression of the mouse retroviruses. Their sequences are aligned (Figure 8-1). A 23-nucleotide region is found to be identical among them and a Blast analysis (NCBI) of these sequences indicates that they are conserved with the v-mos gene, at 3' LTR of the MPSV proviral sequence. They are also conserved with LTR of other MMuLV viruses. The expression of these sequences in AB2.2 cells is consistent with the knowledge of this cell line, in which the *Hprt1* gene was silenced by a recombinant MPSV (King *et al.* 1985; Kuehn *et al.* 1987; Lobel *et al.* 1985).

The other ten genes with ~2 times the expression in the AB1 cell line than in the AB2.2 cell line are not expected. These transcriptional variants could be due to genetic or epigenetic alterations between the two cell lines. From this comparison analysis between two wild type cell lines, the Illumina® expression array platform is confirmed to be a stable platform to measure transcriptional variants between cell lines. It has demonstrated sensitivity to exogenous and endogenous gene expression, and it exhibits that two times the expression level can serve as a threshold to distinguish transcribed from non-transcribed genes. However, it should be borne in mind that a large expression fold change will result from a knock-down of an expressed gene.

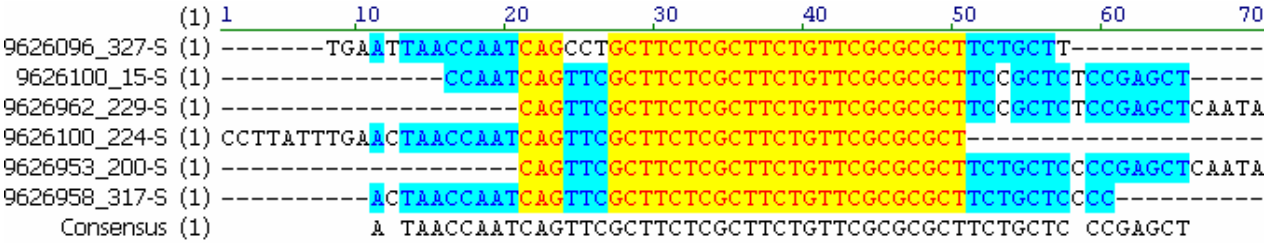


Figure 8-1 Alignment of probe sequences of MPSV

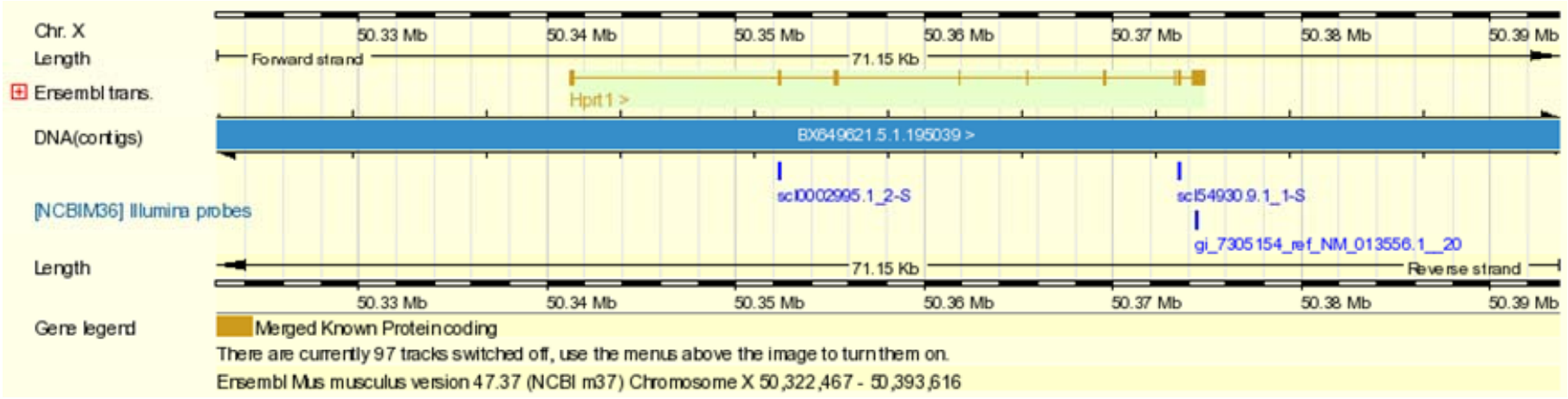
Probe names are listed on the left. These sequences detect transcripts which are expressed twice as much in the AB2.2 compared with the AB1 cell line. Their sequences are aligned and identical nucleotides are shown in a red font on a yellow background. Conserved nucleotides are shown in a dark blue font on a light blue background. Their consensus sequence is shown at the bottom.

Table 8-1 Main transcriptional variants between the AB1 and AB2.2 cell lines

Probe	Gene symbol	Expression fold change	Comparison P-value	Gene description
Expression AB1/AB2.2 ≥ 2.0		**AB1/AB2.2**		
scl30469.7_1-S	*Igf2*	3.89	1.90×10^{-8}	Mus musculus insulin-like growth factor 2 (Igf2), mRNA.
scl34188.7.1_72-S	*Acta1*	3.78	5.60×10^{-9}	Mus musculus actin, alpha 1, skeletal muscle (Acta1), mRNA.
scl54930.9.1_1-S	*Hprt1*	2.75	1.04×10^{-10}	Mus musculus hypoxanthine guanine phosphoribosyl transferase (Hprt), mRNA.
gi_7305154_ref_NM_013556.1__205-S	*Hprt1*	2.73	1.88×10^{-10}	Mus musculus hypoxanthine guanine phosphoribosyl transferase (Hprt), mRNA.
scl0068377.2_268-S		2.71	1.27×10^{-9}	Mus musculus RIKEN cDNA 0610041G09 gene (0610041G09Rik), mRNA.
scl24990.1.20_266-S	*Pou3f1*	2.51	3.08×10^{-13}	Mus musculus POU domain, class 3, transcription factor 1 (Pou3f1), mRNA.
scl54328.4.1_36-S	*Rhox9*	2.43	8.40×10^{-10}	Mus musculus placenta specific homeobox 2 (Psx2), mRNA.
scl16803.9_203-S	*Slc40a1*	2.39	1.57×10^{-9}	Mus musculus solute carrier family 40 (iron-regulated transporter), member 1 (Slc40a1), mRNA.
scl30493.4.19_120-S	*Sct*	2.20	4.52×10^{-8}	Mus musculus secretin (Sct), mRNA.
scl014468.10_101-S	*Gbp1*	2.16	8.60×10^{-11}	Mus musculus guanylate nucleotide binding protein 1 (Gbp1), mRNA.
scl50828.7.1_115-S	*Psmb8*	2.13	7.77×10^{-17}	Mus musculus proteosome (prosome, macropain) subunit, beta type 8 (large multifunctional protease 7) (Psmb8), mRNA.
scl52209.6.1_16-S	*Ttr*	2.08	2.54×10^{-6}	Mus musculus transthyretin (Ttr), mRNA.
scl018715.6_137-S	*Pim2*	2.04	2.70×10^{-13}	Mus musculus proviral integration site 2 (Pim2), mRNA.
scl0002995.1_2-S *	*Hprt1*	1.98	7.99×10^{-33}	Mus musculus hypoxanthine guanine phosphoribosyl transferase (Hprt), mRNA.
Expression AB2.2/AB1 ≥ 2.0		**AB2.2/AB1**		
9626962_229-S #	-	1.92	8.60×10^{-8}	Moloney murine sarcoma virus
9626096_327-S	-	2.01	1.09×10^{-8}	Friend murine leukemia virus
9626953_200-S	-	2.04	1.13×10^{-8}	Abelson murine leukemia virus
9626958_317-S	-	2.04	1.45×10^{-8}	Murine leukemia virus
9626100_15-S	-	2.10	1.07×10^{-10}	Murine sarcoma virus
9626100_224-S	-	2.30	1.46×10^{-11}	Murine sarcoma virus

*: This probe has 1.98 times the signal strength in AB1 than in AB2.2. #: This probe has 1.92 times the signal strength in AB2.2 than in AB2.2.

A



B

Illumina array probe on *Hprt1*	Genomic position on chromosome X (bp)	Within exon	Expression fold change AB1 vs AB2.2	Comparison P-value
scl0002995.1_2-S	49243733~49243782	2	1.98	7.99×10^{-33}
scl54930.9.1_1-S	49264970~49265019	8	2.75	1.04×10^{-10}
gi_7305154_ref_NM_013556.1__205-S	49265893~49265942	9	2.73	1.88×10^{-10}

Figure 8-2 **Mouse *Hprt1* gene expression variation between AB1 and AB2.2 cells**

A. Mouse *Hprt1* gene structure on NCBI m36, Ensembl release 46 with three Illumina® array probes annotated in this gene (blue text). **B**. Summary of expression fold change of these probes between the AB1 and AB2.2 cell lines. Probe coordinates are shown. Results indicate that *Hprt1* transcripts in the AB1 cell line are about 2 times more than those in the AB2.2 cell line. Comparison P-values (adjusted P-values) are near zero (10^{-33}–10^{-10}), indicating that these expression fold changes have a very low probability of being wrong.

8.2.1.2 Comparison between *Blm*-deficient and wild type cells

Since the *Blm*-deficient ES cells are used as the basis of comparisons with the mutant clones, it is important to examine their expression profile against the AB1 and AB2.2 ES cells. Using a comparable analysis strategy, transcriptional variations were analysed between *Blm*-deficient cells (NGG5-3) and wild type cells (AB1 and AB2.2). Data from AB1 and AB2.2 cells, two biological replicates each, was grouped together as the data of wild type cells. There were 111 genes expressing in wild type cells whose expression signals were two times more than those in the *Blm*-deficient cells (Table 8-2). The maximum expression fold change was seen for *Spink3* (serine peptidase inhibitor, Kazal type 3) with 7.4 times the expression in the wild type cells than in *Blm*-deficient cells. On the other hand, there were 67 genes expressing in *Blm*-deficient cells whose expression levels were more than twice of those in wild type cells (Table 8-3). The maximum expression fold change was of the AF067061 gene (2-cell-stage, variable group, member 3), which has 5.0 times more expression in the *Blm*-deficient cells than in wild type cells. The fact that so many genes express differently between the wild type and the *Blm*-deficient cells may be because of two reasons: 1. the loss of the genome surveillance function of the BLM protein. 2. homozygous deletion of transcription factors. All of these genes are distributed throughout the whole genome without a preference for a specific chromosome, which suggests that these changes are not due to a specific structural chromosomal alteration in the cells used for this analysis. There were two targeting experiments and two Cre recombination events conducted on a *Blm*-deficient cell line to generate the cell line of NGG5-3 (Guo *et al.* 2004), which was used in this project. These four steps and the long cell culture history may accumulate mutations and these mutations are likely to lose heterozygosity. When such a homozygous mutation occurs to a transcription factor, the gene expression of a batch of genes can be altered. In the Table 8-2, the expression of the *Tcf7l2* (transcription factor 7-like 2) gene is down-regulated to 19% of that in the wild type cells, indicating this gene does not express. Thus, *Tcf7l2* may contribute to such a non-chromosome-specific gene expression variation.

One change that is expected is in the *Blm* gene itself. The *Blm* gene expression level is detected by the probes scl00067.1_13-S in exon 2 and scl31124.24.1_96-S in exon 22, the last exon (Figure 8-3). As described in the Methods, the *Blm*-deficient cell line NGG5-3 has two different alleles: one has a duplicated exon 3 (allele tm3Brd) and the other one lost exon 2 (allele tm4Brd). Due to a reading frame shift caused by the extra exon 3 in allele tm3Brd, the first premature termination codon (PTC) can be generated in the second exon 3

and following exons. As for allele tm4Brd, a new start codon may be used in exon 3 because the original start codon is in exon 2 and exon 2 is deleted. If transcription starts from the new start codon, the reading frame will also be destroyed and an early premature termination codon can be generated within 30 bases' distance of the new start codon. These PTCs can initialize nonsense-mediated mRNA decay (NMD). The signal of the 3' probe scl31124.24.1_96-S is 1.6 times higher in wild type cells than in *Blm*-deficient cells, which is consistent with the genotypes of these cell lines. This expression fold change is less than two times because the wild type *Blm* gene is either not highly expressed or because NMD in *Blm*-deficient cells is not complete. As for the 5' probe scl00067.1_13-S, no obvious expression difference was observed between the wild type cell lines and the *Blm*-deficient cell line. One reason for this may be due to the fact that the *Blm* transcript is quite long (4491 bp), thus the 5' sequence might not be reverse transcribed very efficiently when using oligo d(T) primers to prepare the samples for expression array.

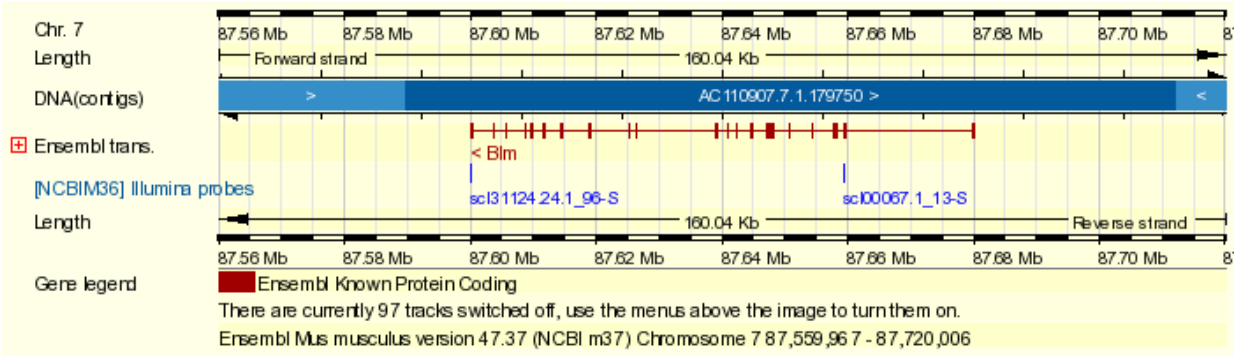


Figure 8-3 Expression array probes in the *Blm* gene

The *Blm* gene is transcribed from the telomere to the centromere direction on the long arm of mouse chromosome 7. There are two probes on the Illumina® array to detect its expression, scl31124.24.1_96-S (80328732–80328781 bp, reverse strand) and scl00067.1_13-S (80388138–80388187 bp, reverse strand).

Table 8-2 **Genes expressed in wild type cells twice as much as in *Blm*-deficient cells**

MGI symbol	Description	% wild type signal	Expression fold change (WT/*Blm*)	Comparison P-value	Chromosome Name
1200002N14Rik	RIKEN cDNA 1200002N14 gene [MGI:1918962]	33%	3.01	4.28×10^{-27}	10
1810015C04Rik	RIKEN cDNA 1810015C04 gene [MGI:1913520]	38%	2.60	4.83×10^{-16}	15
2610042L04Rik	predicted gene, ENSMUSG00000068790 [MGI:3704420]	34%	2.95	2.05×10^{-8}	14
3110043J09Rik	RIKEN cDNA 3110043J09 gene [MGI:1920417]	44%	2.25	1.58×10^{-14}	15
4631426J05Rik	RIKEN cDNA 4631426J05 gene [MGI:1924840]	43%	2.31	2.74×10^{-24}	7
5730469M10Rik	RIKEN cDNA 5730469M10 gene [MGI:1917814]	38%	2.66	7.48×10^{-16}	14
9930013L23Rik	RIKEN cDNA 9930013L23 gene [MGI:2443629]	47%	2.13	1.18×10^{-15}	7
Adam19	a disintegrin and metallopeptidase domain 19 (meltrin beta) [MGI:105377]	49%	2.04	1.27×10^{-19}	11
Amn	amnionless [MGI:1934943]	49%	2.04	9.24×10^{-19}	12
Arhgap17	Rho GTPase activating protein 17 [MGI:1917747]	23%	4.38	4.72×10^{-20}	7
Atp1b1	ATPase, Na+/K+ transporting, beta 1 polypeptide [MGI:88108]	40%	2.48	2.39×10^{-20}	1
Atp6v0a1	ATPase, H+ transporting, lysosomal V0 subunit A1 [MGI:103286]	45%	2.23	3.51×10^{-17}	11
Bach1	BTB and CNC homology 1 [MGI:894680]	49%	2.03	7.17×10^{-18}	16
Bat5	HLA-B associated transcript 5 [MGI:99476]	47%	2.14	5.45×10^{-9}	17
Car12	carbonic anyhydrase 12 [MGI:1923709]	47%	2.14	1.99×10^{-20}	9
Car4	carbonic anhydrase 4 [MGI:1096574]	24%	4.14	1.73×10^{-21}	11
Ccnt2	cyclin T2 [MGI:1920199]	42%	2.38	3.61×10^{-10}	1
Cited2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2 [MGI:1306784]	50%	2.01	3.59×10^{-6}	10
Clic6	chloride intracellular channel 6 [MGI:2146607]	33%	3.01	2.42×10^{-19}	16
Clu	clusterin [MGI:88423]	22%	4.50	1.84×10^{-27}	14
Col4a1	procollagen, type IV, alpha 1 [MGI:88454]	40%	2.53	1.72×10^{-17}	8
Col4a2	procollagen, type IV, alpha 2 [MGI:88455]	24%	4.20	2.10×10^{-26}	8
Csf2ra	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage) [MGI:1339754]	25%	3.94	3.12×10^{-29}	19
Ctgf	connective tissue growth factor [MGI:95537]	37%	2.68	3.12×10^{-29}	10
Ctns	cystinosis, nephropathic [MGI:1932872]	44%	2.27	4.74×10^{-12}	11
Ctsc	cathepsin C [MGI:109553]	49%	2.03	1.03×10^{-19}	7
Ctsh	cathepsin H [MGI:107285]	37%	2.73	4.85×10^{-21}	9
Cubn	cubilin (intrinsic factor-cobalamin receptor) [MGI:1931256]	26%	3.86	2.20×10^{-30}	2

Dab2	disabled homolog 2 (*Drosophila*) [MGI:109175]	44%	2.25	2.27×10^{-16}	15
Dnmt3b	DNA methyltransferase 3B [MGI:1261819]	20%	4.99	1.87×10^{-24}	2
Drp2	dystrophin related protein 2 [MGI:107432]	48%	2.07	6.64×10^{-25}	X
EG245190	predicted gene, EG245190 [MGI:3646425]	34%	2.97	3.54×10^{-11}	7
F3	coagulation factor III [MGI:88381]	35%	2.83	1.84×10^{-27}	3
Fabp3	fatty acid binding protein 3, muscle and heart [MGI:95476]	39%	2.58	1.69×10^{-22}	4
Fhl1	four and a half LIM domains 1 [MGI:1298387]	44%	2.25	7.17×10^{-18}	X
Flrt3	fibronectin leucine rich transmembrane protein 3 [MGI:1918686]	23%	4.26	2.59×10^{-21}	2
Foxq1	forkhead box Q1 [MGI:1298228]	42%	2.36	6.52×10^{-12}	13
Fst	follistatin [MGI:95586]	42%	2.36	2.73×10^{-16}	13
Fv1	Friend virus susceptibility 1 [MGI:95595]	31%	3.25	1.87×10^{-26}	4
Fxyd3	FXYD domain-containing ion transport regulator 3 [MGI:107497]	20%	4.89	1.43×10^{-25}	7
Gata4	GATA binding protein 4 [MGI:95664]	43%	2.33	8.03×10^{-16}	14
Gata6	GATA binding protein 6 [MGI:107516]	42%	2.39	5.91×10^{-12}	18
Gbp1	guanylate nucleotide binding protein 1 [MGI:95666]	32%	3.16	1.01×10^{-29}	3
Gcnt1	glucosaminyl (N-acetyl) transferase 1, core 2 [MGI:95676]	36%	2.81	3.68×10^{-26}	19
Gdpd5	glycerophosphodiester phosphodiesterase domain containing 5 [MGI:2686926]	21%	4.82	4.65×10^{-25}	7
Glipr1	GLI pathogenesis-related 1 (glioma) [MGI:1920940]	44%	2.27	7.78×10^{-11}	10
Gpc3	glypican 3 [MGI:104903]	32%	3.10	9.24×10^{-24}	X
Gpx3	glutathione peroxidase 3 [MGI:105102]	50%	2.01	3.03×10^{-19}	11
Grb10	growth factor receptor bound protein 10 [MGI:103232]	34%	2.93	4.60×10^{-19}	11
Gtl2	GTL2, imprinted maternally expressed untranslated mRNA [MGI:1202886]	35%	2.87	1.45×10^{-18}	12
H19	H19 fetal liver mRNA [MGI:95891]	38%	2.60	1.11×10^{-13}	7
Hkdc1	hexokinase domain containing 1 [MGI:2384910]	26%	3.78	7.73×10^{-30}	10
Hs3st1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1 [MGI:1201606]	25%	3.94	5.00×10^{-18}	5
Kdelr3	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 3 [MGI:2145953]	22%	4.50	5.66×10^{-26}	15
Kit	kit oncogene [MGI:96677]	26%	3.92	2.16×10^{-35}	5
Klb	klotho beta [MGI:1932466]	15%	6.68	9.50×10^{-31}	5
Krt18	keratin 18 [MGI:96692]	30%	3.32	1.39×10^{-26}	15
Krt8	keratin 8 [MGI:96705]	41%	2.46	2.43×10^{-16}	15
Lamb1-1	laminin B1 subunit 1 [MGI:96743]	41%	2.43	5.72×10^{-20}	12
Lamc1	laminin, gamma 1 [MGI:99914]	34%	2.93	9.81×10^{-10}	1
Ldlrap1	low density lipoprotein receptor adaptor protein 1 [MGI:2140175]	46%	2.19	1.86×10^{-6}	4

Lgmn	legumain [MGI:1330838]	41%	2.46	8.76×10^{-13}	12
Loxl2	lysyl oxidase-like 2 [MGI:2137913]	24%	4.23	5.22×10^{-35}	14
Lrp2	low density lipoprotein receptor-related protein 2 [MGI:95794]	38%	2.64	8.19×10^{-26}	2
Lrpap1	low density lipoprotein receptor-related protein associated protein 1 [MGI:96829]	27%	3.73	3.68×10^{-26}	5
Luc7l2	LUC7-like 2 (S. cerevisiae) [MGI:2183260]	34%	2.97	7.65×10^{-25}	6
Mod1	malic enzyme, supernatant [MGI:97043]	25%	4.03	3.17×10^{-19}	9
Myl3	myosin, light polypeptide 3 [MGI:97268]	48%	2.07	1.54×10^{-11}	9
Nav1	neuron navigator 1 [MGI:2183683]	46%	2.16	4.14×10^{-6}	1
Nostrin	nitric oxide synthase trafficker [MGI:3606242]	35%	2.87	3.99×10^{-16}	2
Nrk	Nik related kinase [MGI:1351326]	27%	3.73	9.31×10^{-22}	X
Nxf7	nuclear RNA export factor 7 [MGI:2159343]	48%	2.10	4.75×10^{-17}	X
Palld	palladin, cytoskeletal associated protein [MGI:1919583]	32%	3.10	1.97×10^{-17}	8
Pcgf5	polycomb group ring finger 5 [MGI:1923505]	45%	2.22	5.24×10^{-16}	19
Pdcd4	programmed cell death 4 [MGI:107490]	26%	3.84	1.01×10^{-29}	19
Pdpn	podoplanin [MGI:103098]	45%	2.22	1.03×10^{-18}	4
Peg3	paternally expressed 3 [MGI:104748]	46%	2.16	2.10×10^{-17}	7
Phf17	PHD finger protein 17 [MGI:1925835]	47%	2.14	8.03×10^{-14}	3
Pla2g12b	phospholipase A2, group XIIB [MGI:1917086]	49%	2.06	7.11×10^{-16}	10
Plod2	procollagen lysine, 2-oxoglutarate 5-dioxygenase 2 [MGI:1347007]	46%	2.17	3.44×10^{-12}	9
Podxl	podocalyxin-like [MGI:1351317]	42%	2.36	2.84×10^{-19}	6
Pthr1	parathyroid hormone receptor 1 [MGI:97801]	39%	2.57	1.67×10^{-26}	1
Pthr1	parathyroid hormone receptor 1 [MGI:97801]	44%	2.28	1.39×10^{-13}	9
Ptk6	Ptk6 protein tyrosine kinase 6 [MGI:99683]	27%	3.71	3.59×10^{-21}	2
Rab15	RAB15, member RAS oncogene family [MGI:1916865]	41%	2.43	1.06×10^{-19}	12
Ramp2	receptor (calcitonin) activity modifying protein 2 [MGI:1859650]	48%	2.08	5.03×10^{-25}	11
Rasd1	RAS, dexamethasone-induced 1 [MGI:1270848]	40%	2.53	6.37×10^{-16}	11
Rassf3	Ras association (RalGDS/AF-6) domain family 3 [MGI:2179722]	42%	2.36	6.27×10^{-19}	10
Rhox9	reproductive homeobox 9 [MGI:1890128]	39%	2.57	2.66×10^{-27}	X
S100a1	S100 calcium binding protein A1 [MGI:1338917]	41%	2.43	1.78×10^{-19}	3
Serpinh1	serine (or cysteine) peptidase inhibitor, clade H, member 1 [MGI:88283]	46%	2.17	9.00×10^{-19}	7
Sfmbt2	Scm-like with four mbt domains 2 [MGI:2447794]	45%	2.23	2.31×10^{-8}	2
Slc38a2	solute carrier family 38, member 2 [MGI:1915010]	44%	2.28	1.25×10^{-8}	15
Slc40a1	solute carrier family 40 (iron-regulated transporter), member 1 [MGI:1315204]	45%	2.20	1.47×10^{-13}	1
Slco2a1	solute carrier organic anion transporter family, member 2a1 [MGI:1346021]	33%	3.01	2.48×10^{-28}	9

Soat1	sterol O-acyltransferase 1 [MGI:104665]	40%	2.51	2.87×10^{-18}	1
Sox17	SRY-box containing gene 17 [MGI:107543]	42%	2.36	1.70×10^{-12}	1
Sox7	SRY-box containing gene 7 [MGI:98369]	41%	2.41	1.45×10^{-9}	14
Spink3	serine peptidase inhibitor, Kazal type 3 [MGI:106202]	29%	3.41	2.98×10^{-21}	18
Srgn	serglycin [MGI:97756]	47%	2.14	7.46×10^{-16}	10
Stard8	START domain containing 8 [MGI:2448556]	16%	6.15	5.05×10^{-29}	X
Syvn1	synovial apoptosis inhibitor 1, synoviolin [MGI:1921376]	32%	3.14	2.65×10^{-27}	19
Tax1bp3	Tax1 (human T-cell leukemia virus type I) binding protein 3 [MGI:1923531]	13%	7.41	4.40×10^{-20}	11
Tcf7l2	transcription factor 7-like 2, T-cell specific, HMG-box [MGI:1202879]	19%	5.31	3.51×10^{-17}	19
Tfpi	tissue factor pathway inhibitor [MGI:1095418]	37%	2.71	5.45×10^{-21}	2
Tinagl	tubulointerstitial nephritis antigen-like [MGI:2137617]	47%	2.11	3.31×10^{-22}	4
Tmem166	transmembrane protein 166 [MGI:2385247]	39%	2.55	7.23×10^{-25}	6
Tnk1	tyrosine kinase, non-receptor, 1 [MGI:1930958]	42%	2.38	1.85×10^{-24}	11
Tspan2	tetraspanin 2 [MGI:1917997]	43%	2.33	2.32×10^{-9}	3
Ttr	transthyretin [MGI:98865]	37%	2.73	5.40×10^{-16}	18
Txndc12	thioredoxin domain containing 12 (endoplasmic reticulum) [MGI:1913323]	39%	2.58	2.98×10^{-21}	4

Table 8-3 **Genes expressed in *Blm*-deficient cells twice as much as in wild type cells**

MGI symbol	Description	Expression fold change (*Blm*/WT)	Comparison P-value	Chromosome Name
2200001I15Rik	RIKEN cDNA 2200001I15 gene [MGI:1916384]	2.31	4.51×10^{-17}	14
2310016C16Rik	RIKEN cDNA 2310016C16 gene [MGI:1916840]	2.97	1.58×10^{-11}	13
2310039H08Rik	RIKEN cDNA 2310039H08 gene [MGI:1914351]	2.08	4.88×10^{-15}	17
2410116G06Rik	RIKEN cDNA 2410116G06 gene [MGI:1915486]	2.30	3.07×10^{-11}	2
6330406I15Rik	RIKEN cDNA 6330406I15 gene [MGI:1917967]	2.99	2.98×10^{-9}	5
Aebp1	AE binding protein 1 [MGI:1197012]	2.04	1.60×10^{-5}	11
AF067061	2-cell-stage, variable group, member 3 [MGI:2675349]	4.99	5.74×10^{-16}	NT_053651
Antxr2	anthrax toxin receptor 2 [MGI:1919164]	2.13	6.00×10^{-10}	5
Aqp1	aquaporin 1 [MGI:103201]	2.20	2.65×10^{-8}	6
Axl	AXL receptor tyrosine kinase [MGI:1347244]	2.55	2.97×10^{-10}	7
Bgn	biglycan [MGI:88158]	2.23	1.49×10^{-9}	X
Cav1	caveolin, caveolae protein 1 [MGI:102709]	2.23	3.50×10^{-9}	6
Cd44	CD44 antigen [MGI:88338]	2.10	2.93×10^{-9}	2
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	2.43	6.22×10^{-9}	4
Col4a5	procollagen, type IV, alpha 5 [MGI:88456]	2.35	2.18×10^{-11}	X
Col6a1	procollagen, type VI, alpha 1 [MGI:88459]	2.62	1.36×10^{-8}	10
Cpe	carboxypeptidase E [MGI:101932]	2.14	7.82×10^{-8}	8
Crlf1	cytokine receptor-like factor 1 [MGI:1340030]	2.19	5.29×10^{-12}	8
Cul1	cullin 1 [MGI:1349658]	2.08	3.44×10^{-12}	6
Cxcl1	chemokine (C-X-C motif) ligand 1 [MGI:108068]	2.03	3.35×10^{-10}	5
D12Ertd647e	ISG12a protein. [Source:Uniprot/SPTREMBL;Acc:Q70LN0]	2.36	2.89×10^{-9}	12
Dcn	decorin [MGI:94872]	3.27	3.56×10^{-8}	10
Dlk1	delta-like 1 homolog (*Drosophila*) [MGI:94900]	2.16	1.64×10^{-4}	12
Ecm1	extracellular matrix protein 1 [MGI:103060]	2.27	8.73×10^{-12}	3
Emp3	epithelial membrane protein 3 [MGI:1098729]	2.46	6.29×10^{-9}	7
Eno2	enolase 2, gamma neuronal [MGI:95394]	2.10	1.15×10^{-10}	6
Fbln2	fibulin 2 [MGI:95488]	2.10	3.08×10^{-7}	6
Fez1	fasciculation and elongation protein zeta 1 (zygin I) [MGI:2670976]	2.07	1.41×10^{-16}	9
Fxyd5	FXYD domain-containing ion transport regulator 5 [MGI:1201785]	2.20	1.97×10^{-11}	7
Gbx2	gastrulation brain homeobox 2 [MGI:95668]	2.73	2.75×10^{-12}	1

H2-D1	histocompatibility 2, D region [MGI:95912]	3.07	2.06×10^{-13}	17
Ifitm3	interferon induced transmembrane protein 3 [MGI:1913391]	2.50	2.97×10^{-15}	7
Igfbp7	insulin-like growth factor binding protein 7 [MGI:1352480]	2.16	2.07×10^{-10}	5
Il1rl1	interleukin 1 receptor-like 1 [MGI:98427]	2.66	8.63×10^{-8}	1
Jam2	junction adhesion molecule 2 [MGI:1933820]	2.57	5.06×10^{-23}	16
Lgals1	lectin, galactose binding, soluble 1 [MGI:96777]	2.83	9.08×10^{-15}	15
Lmna	lamin A [MGI:96794]	2.20	4.23×10^{-10}	3
Lox	lysyl oxidase [MGI:96817]	2.11	1.94×10^{-9}	18
Loxl1	lysyl oxidase-like 1 [MGI:106096]	2.62	4.70×10^{-14}	9
Ltbp2	latent transforming growth factor beta binding protein 2 [MGI:99502]	2.53	6.86×10^{-10}	12
Ltbp3	latent transforming growth factor beta binding protein 3 [MGI:1101355]	2.71	4.64×10^{-11}	19
Ly6a	lymphocyte antigen 6 complex, locus A [MGI:107527]	2.45	1.71×10^{-9}	15
Masp1	mannan-binding lectin serine peptidase 1 [MGI:88492]	2.08	1.75×10^{-8}	16
Mmd	monocyte to macrophage differentiation-associated [MGI:1914718]	2.31	8.32×10^{-11}	11
Mmp14	matrix metallopeptidase 14 (membrane-inserted) [MGI:101900]	2.14	1.07×10^{-11}	14
Msn	moesin [MGI:97167]	2.00	8.45×10^{-12}	X
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle [MGI:97273]	3.51	4.15×10^{-18}	7
Myo1f	myosin IF [MGI:107711]	2.30	4.40×10^{-20}	17
Notch4	Notch gene homolog 4 (*Drosophila*) [MGI:107471]	2.01	2.22×10^{-13}	17
Nrp1	neuropilin 1 [MGI:106206]	2.03	3.14×10^{-10}	8
Nupr1	nuclear protein 1 [MGI:1891834]	2.13	8.46×10^{-18}	7
Plau	plasminogen activator, urokinase [MGI:97611]	2.69	7.03×10^{-9}	14
Prkg2	protein kinase, cGMP-dependent, type II [MGI:108173]	2.31	8.04×10^{9}	5
Ptn	pleiotrophin [MGI:97804]	2.25	3.01×10^{-9}	6
Rbp1	retinol binding protein 1, cellular [MGI:97876]	2.13	4.20×10^{-11}	9
S100a4	S100 calcium binding protein A4 [MGI:1330282]	2.48	3.40×10^{-9}	3
Sdpr	serum deprivation response [MGI:99513]	2.01	4.37×10^{-11}	1
Serpinf1	serine (or cysteine) peptidase inhibitor, clade F, member 1 [MGI:108080]	2.27	4.38×10^{-7}	11
Spon2	spondin 2, extracellular matrix protein [MGI:1923724]	2.10	1.98×10^{-8}	5
Tcf15	transcription factor 15 [MGI:104664]	2.25	4.44×10^{-14}	2
Tcstv3	2-cell-stage, variable group, member 3 [MGI:2675349]	3.32	1.71×10^{-13}	NT_053651
Tnfrsf22	tumour necrosis factor receptor superfamily, member 22 [MGI:1930270]	2.41	3.83×10^{-12}	7
Tnnt2	troponin T2, cardiac [MGI:104597]	2.31	9.78×10^{-12}	1
Tspo	translocator protein [MGI:88222]	2.08	4.03×10^{-11}	15

Ube1l2	ubiquitin-activating enzyme E1-like 2 [MGI:1913894]	2.60	5.81×10^{-12}	5
Ugt1a6b	UDP glucuronosyltransferase 1 family, polypeptide A7C [MGI:3032636]	2.14	1.20×10^{-7}	1
Vcam1	vascular cell adhesion molecule 1 [MGI:98926]	2.27	3.37×10^{-10}	3

8.2.1.3 Comparison between *Dnmt1*-deficient and AB2.2 wild type cells

The *Dnmt1* gene is involved in DNA MMR (Guo *et al.* 2004). Its deficiency leads to 6TG resistance and microsatellite instability. One expected change from this comparison is the expression change in the *Dnmt1* gene. This gene has one probe (scl36157.39.1_18-S) on the expression array. However, the expression signal of it is 76% of that in the AB2.2 wild type cells. This *Dnmt1*-deficient cell line was generated by deleting the exons 2–4 thourgh gene targeting and a frame shift mutation was generated in this gene. Loss of function of DNMT1 protein is resulted from this frame shift mutation while its gene transcripts can still exist if nonsense-mediated decay does not degrade its transcripts efficiently. The scl36157.39.1_18-S probe locates in the exon 39, the last exon of the *Dnmt1* gene (Figure 8-11), thus its signal may not be affected.

To explore the downstream regulated genes in *Dnmt1*-deficient cells, the expression profile of the ww56+*Hprt* cell line was compared with that of the AB2.2 cell line, from which ww56+*Hprt* was derived. It is assumed that some genes regulated by *Dnmt1* may also be regulated by other MMR components. Thus, the comparison data between the ww56+*Hprt* and AB2.2 cell lines provides a reference for identifying MMR proteins regulated genes in 6TG-resistant mutants.

The expression signals of ten genes in the *Dnmt1*-deficient cell line were two or more times greater than those of the AB2.2 wild type cell line (Table 8-4). The maximum expression fold difference is 3.5 fold for the *Fv1* gene. In addition, the signal of the *neomycin* gene probe is 16 times higher in the *Dnmt1*-deficient cell line ww56+*Hprt*. This is consistent with the selection marker used in the ww56+*Hprt* cell line. Three genes (*2410116G06Rik*, *Chac1* and *Mybl2*) on chromosome 2 and three genes (*Rhox2*, *Scml2* and *Xlr3a*) on chromosome X show altered expression, however, these genes are far from each other. Twenty-two genes are expressed more than two times higher in AB2.2 wild type ES cells than in *Dnmt1*-deficient cells (Table 8-5). The maximum expression fold difference was 9.2 fold for the *Eif2s3y* gene. There is only one probe on the Illumina® array for the *Dnmt1* gene. This probe, scl36157.39.1_18-S, is the same as the genomic sequence between 20657646 and 20657695 bp. It is in the 3' end of the gene on the reverse strand of chromosome 9. The signal of this probe is 1.3 times higher in the AB2.2 than in the ww56+*Hprt* cell line, similar to the *Blm* gene probe in the *Blm*-deficient cells, this modest expression difference may be due to uncompleted NMD.

Table 8-4 **Genes expressed more than two fold higher in the *Dnmt1*-deficient cells compared with the AB2.2 wild type cells**

MGI symbol	Description	Expression fold change (Dnmt1/AB2.2)	Comparison P-value	Chromosome Name	Gene Start (bp)	Gene End (bp)
2410116G06Rik	RIKEN cDNA 2410116G06 gene [MGI:1915486]	2.14	3.66×10^{-11}	2	162912770	162913591
Chac1	ChaC, cation transport regulator-like 1 (E. coli) [MGI:1916315]	2.57	3.70×10^{-9}	2	119176992	119180117
Fv1	Friend virus susceptibility 1 [MGI:95595]	3.51	1.73×10^{-15}	4	147243088	147244467
Gbx2	gastrulation brain homeobox 2 [MGI:95668]	2.14	8.24×10^{-6}	1	91824559	91827751
Mybl2	myeloblastosis oncogene-like 2 [MGI:101785]	2.03	4.74×10^{-10}	2	162880371	162910423
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle [MGI:97273]	2.75	2.47×10^{-10}	7	134355122	134357801
Rhox2	reproductive homeobox 2 [MGI:1922449]	2.06	2.23×10^{-20}	X	34784987	34789685
Scml2	sex comb on midleg-like 2 (*Drosophila*) [MGI:1340042]	2.01	2.02×10^{-12}	X	157555125	157696145
Sgip1	SH3-domain GRB2-like (endophilin) interacting protein 1 [MGI:1920344]	2.07	7.32×10^{-09}	4	102413902	102644468
Xlr3a	X-linked lymphocyte-regulated 3A [MGI:109506]	2.17	3.84×10^{-12}	X	70331634	70342380

These ten genes increased their expression in the *Dnmt1*-deficient cells to 2–3.5 times of those in the parental cell line, AB2.2.

Table 8-5 **Genes expressed more than two fold higher in the AB2.2 wild type cells compared with the *Dnmt1*-deficient cells**

MGI symbol	Description	% wild type signal	Expression fold change (AB2.2/*Dnmt1*)	Comparison P-value	Chromosome Name
Amn	amnionless [MGI:1934943]	46%	2.16	2.77×10^{-6}	12
Car12	carbonic anyhydrase 12 [MGI:1923709]	49%	2.04	1.14×10^{-6}	9
Cdkn1c	cyclin-dependent kinase inhibitor 1C (P57) [MGI:104564]	44%	2.25	5.73×10^{-9}	7
Clic6	chloride intracellular channel 6 [MGI:2146607]	45%	2.22	1.16×10^{-11}	16
Clu	clusterin [MGI:88423]	30%	3.39	3.89×10^{-18}	14
Col4a2	procollagen, type IV, alpha 2 [MGI:88455]	49%	2.03	2.42×10^{-11}	8
Ctsh	cathepsin H [MGI:107285]	40%	2.51	3.63×10^{-17}	9
EG245190	predicted gene, EG245190 [MGI:3646425]	45%	2.20	1.57×10^{-20}	7
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked [MGI:1349430]	11%	9.19	1.63×10^{-33}	Y
F3	coagulation factor III [MGI:88381]	43%	2.31	3.76×10^{-8}	3
Fst	follistatin [MGI:95586]	47%	2.14	3.81×10^{-10}	13
Fxyd3	FXYD domain-containing ion transport regulator 3 [MGI:107497]	44%	2.27	2.53×10^{-17}	7
Grb10	growth factor receptor bound protein 10 [MGI:103232]	23%	4.29	2.58×10^{-27}	11
Hkdc1	hexokinase domain containing 1 [MGI:2384910]	40%	2.53	3.33×10^{-23}	10
Hs3st1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1 [MGI:1201606]	44%	2.30	4.22×10^{-10}	5
Igf2	insulin-like growth factor 2 [MGI:96434]	14%	6.96	5.88×10^{-16}	7
Lamc1	laminin, gamma 1 [MGI:99914]	47%	2.13	2.19×10^{-16}	1
Phlda2	pleckstrin homology-like domain, family A, member 2 [MGI:1202307]	38%	2.62	2.43×10^{-12}	7
Podxl	podocalyxin-like [MGI:1351317]	46%	2.16	2.72×10^{-8}	6
Slco2a1	solute carrier organic anion transporter family, member 2a1 [MGI:1346021]	49%	2.04	4.34×10^{-16}	9
Spink3	serine peptidase inhibitor, 0Kazal type 3 [MGI:106202]	39%	2.55	2.69×10^{-5}	18

The above genes have decreased their expression to 10–50% of that in the AB2.2 wild type cells.

8.2.1.4 *Msh2* and *Msh6* homozygous mutants

There are five mutants (B6, D4, D8, F4 and H14), in which both *Msh2* and *Msh6* genes are homozygously deleted. All of the mutant clones with *Msh2* and *Msh6* deletions were analysed by expression arrays to validate the array for deletions of the homozygously deleted genes and to obtain a downstream regulated gene profile of these MMR mutants. The downstream genes are possibly regulated by MMR defects, thus consistent patterns may be observed in other MMR mutants. A gene regulatory profile can also be used as a reference to judge if another mutant has an *Msh2* and *Msh6* deficiency. In the analysis, I expected to see that the *Msh2* and *Msh6* genes were not expressed.

On the Illumina® mouse expression array, *Msh2* has three probes: ri|8030473G08|PX00650P10|AK078, ri|D230024H20|PX00188O22|AK051 and scl50397.16.1_74-S (Figure 8-4). Among them, probe scl50397.16.1_74-S gave the most reliable data since its expression was more than four times higher in the *Blm*-deficient ES cells than in the mutants. The expression signal of probe ri|8030473G08|PX00650P10|AK078 was only 1.38 times higher in the *Blm*-deficient cell line NGG5-3 than in the mutants, which may be regarded as non-complementary cross-hybridization, as this level of expression difference was also observed between wild type cells. The signal of ri|D230024H20|PX00188O22|AK051 was similar to background, thus it was not counted.

Msh6 has two probes on the array: scl50396.10.5_23-S and scl0017688.1_95-S positioed in exons 9 and 10, respectively (Figure 8-5). The expression signal of probe scl0017688.1_95-S was 30–47 times higher in the *Blm*-deficient cells than in the five clones which have *Msh2* and *Msh6* deleted. The signal of the other probe scl50396.10.5_23-S was only 1.4–1.5 times higher in the *Blm*-deficient cells compared with the mutants, indicating that the signal detected by this probe may come from non-complementary hybridization.

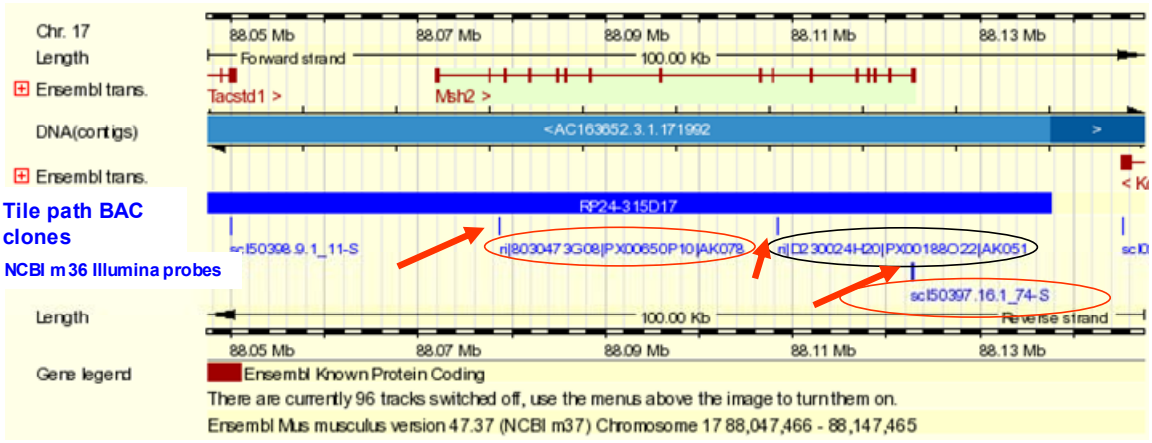
These analysis indicated that *Msh2* and *Msh6* are not expressed as expected. However, this analysis also illustrated that many of the probes on the Illumina array were not that reliable. This can also be observed from the previous data that the probes of *Blm*, *Dnmt1* and *Hprt* genes only show 1–2 times signal difference or no difference between the wild type cells and the gene deficient cells. The genes which show a change in expression are listed in

Table 8-6. The array profiles of these mutants provide an opportunity to determine if there are any other consistent transcriptional differences between the mutants. By comparing all the mutants, 15 genes were found to be down-regulated more than 50% in all of the *Msh2* and *Msh6* homozygous mutants. These genes are distributed throughout in the mouse genome thus are not linked with the homozygous deletion of chromosome 17, though *Msh2* and *Msh6* are both detected by this analysis. Although no evidence has been found for any of these genes to be involved in DNA mismatch repair until now, it can be suggested that these genes are downstream indirectly associated with either *Msh2* function or *Msh6* function. However, other possibilities should be borne in mind, such as that these expression variations could be permanently introduced by gamma irradiation. The genes in the common homozygous deleted region were expected to be found in the Table 8-6 but they were not. These genes have probes on the expression array. Their signals might be excluded from the analysis due to relatively big variations between replicates.

By comparing all the mutants to detect upregulated genes, only one probe scl0012865.1_86-S was detected which showed 2.0–3.5 times increased expression. This probe was not annotated correctly in the mouse genome, but Blast analysis identified this as part of the *Cox7a1* transcript (Figure 8-6). Interestingly, the expression signal of this gene in the *Dnmt1* knockout cells (ww56+*Hprt*) is 1.48 times more than that in the AB2.2 wild type cells. The *Cox7a1* gene is the mouse homologue of the cytochrome c oxidase, subunit VIIa 1 (*Cox7a1*) in humans. COX7A1 is a nuclear-encoded subunit, which may play a regulatory role in the fuction of cytochrome c oxidase (OMIM). The cytochrome c oxidase is the last component of the mitochondrial respiratory chain. This oxidase catalyzes the transfer of electrons from reduced cytochrome c to molecular oxygen. No relation between this gene and MMR was found in literature. It could be related to MMR-deficiency. To elucidate this, more expression analysis on targeted or gene trapped $Msh2^{-/-}$ or $Msh6^{-/-}$ cells with normal BLM function is required.

Based on the above analysis, the expression signals of individual probes were consistent in these clones. This helps to validate the performance of the probes on the other arrays. If the high quality data from these probes can be extended to all of the probes, whole genome expression profiles can be compared confidently.

A



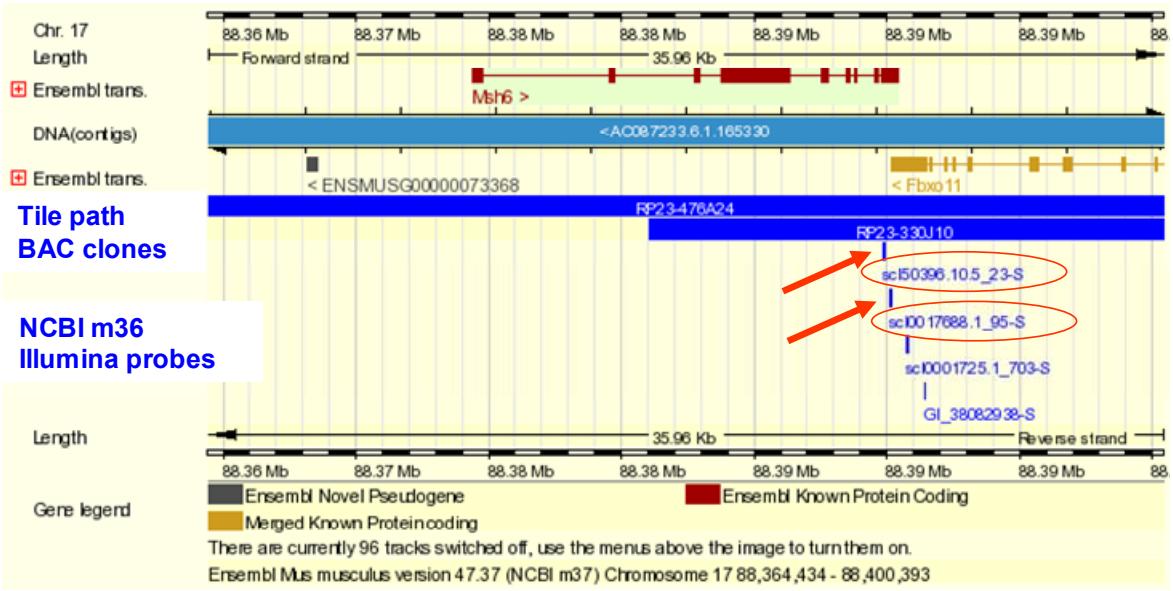
B

Clone	Expression fold change (*Blm*/mutant)		Expression change (mutant/*Blm*)	
	ri|8030473G08|PX00650P10|AK078	scl50397.16.1_74-S	ri|8030473G08|PX00650P10|AK078	scl50397.16.1_74-S
B6	1.38	4.32	72%	23%
D4	1.36	4.63	74%	22%
D8	1.38	4.79	72%	21%
F4	1.39	4.47	72%	22%
H14	1.39	4.56	72%	22%

Figure 8-4 Expression fold change of the *Msh2* gene in homozygously deleted mutants

A. The expression array probes are annotated (small blue bars) on the *Msh2* transcript (NCBI m37, Ensembl release 47). The *Msh2* gene has three probes (circled): ri|8030473G08|PX00650P10|AK078, ri|D230024H20|PX00188O22|AK051 and scl50397.16.1_74-S. Their genomic positions are indicated by red arrows. The genomic DNA of BAC RP24-315D17 is shown in a large blue bar. This BAC was used in the 200 kb resolution tile path array CGH. **B**. The expression level of *Msh2* probes compared to the *Blm*-deficient cells. The expression signal of probe scl50397.16.1_74-S (red circled) in the *Blm*-deficient cell line is around four times more than that in the mutants. The expression signal of probe ri|8030473G08|PX00650P10|AK078 (red circled) in the *Blm*-deficient cells is around 1.38 times more than that in the mutants, suggesting that the signals are probably due to non-complementary cross-hybridization. The signal of probe ri|D230024H20|PX00188O22|AK051 (black circled) was not consistent among its features on the arrays, thus it was not counted.

A



B

Clone	Expression fold change (*Blm*/mutant)		Expression change (mutant/*Blm*)	
	scl50396.10.5_23-S	scl0017688.1_95-S	scl50396.10.5_23-S	scl0017688.1_95-S
B6	1.48	30.7	68%	3%
D1	1.48	42.8	68%	2%
D4	1.49	43.4	67%	2%
D8	1.46	38.9	68%	3%
F4	1.51	36.3	66%	3%
H14	1.49	46.9	67%	2%

Figure 8-5 Expression fold change of the *Msh6* gene in homozygously deleted mutants

A. The expression array probes are annotated (small blue bars) on the *Msh6* transcript (NCBI m37, Ensembl release 47). The *Msh6* gene has two probes (red circled, their positions are indicated by red arrows): scl50396.10.5_23-S (on exon 9) and scl0017688.1_95-S (on exon 10, the last exon). The genomic DNA of BAC RP23-476A24 and RP23-330J10 is shown in large blue bars. These BACs were used in the 200 kb resolution tile path array CGH. **B**. The expression level of *Msh6* probes compared to the *Blm*-deficient cells. The expression signal of probe scl0017688.1_95-S in the *Blm*-deficient cells is around 30–47 times more than that in the mutants. The signal of the other probe, scl50396.10.5_23-S, in the *Blm*-deficient cells is around 1.4–1.5 times more than that in the mutants, indicating that the signal detected by this probe may come from non-complementary hybridization.

Table 8-6 **Down-regulated genes in all five *Msh2* and *Msh6* homozygously deleted mutants compared with the *Blm*-deficient NGG5-3 cell line**

MGI symbol	Description	% $Blm^{-/-}$ signal					Average % $Blm^{-/-}$ signal ± SD	Chromosome Name	Gene Start (bp)	Gene End (bp)
		B6	D4	D8	F4	H14				
2310016C16Rik	RIKEN cDNA 2310016C16 gene [MGI:1916840]	44%	28%	31%	37%	31%	34±6%	13	113832971	113836596
Ank	progressive ankylosis [MGI:3045421]	41%	35%	42%	38%	39%	39±3%	15	27396432	27524660
Bst2	bone marrow stromal cell antigen 2 [MGI:1916800]	19%	27%	42%	34%	34%	31±9%	8	74058163	74061336
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	47%	38%	40%	46%	35%	41±5%	4	129725129	129776519
Col6a1	procollagen, type VI, alpha 1 [MGI:88459]	50%	42%	44%	45%	29%	42±8%	10	76171537	76188789
Cpe	carboxypeptidase E [MGI:101932]	50%	40%	40%	46%	38%	43±5%	8	67071355	67171851
Dcn	decorin [MGI:94872]	33%	22%	27%	25%	28%	27±4%	10	96942245	96980777
Ecm1	Extracellular matrix protein 1 [MGI:103060]	39%	47%	48%	44%	43%	44±4%	3	95538073	95543494
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked [MGI:1349430]	12%	12%	12%	11%	13%	12±1%	Y	347055	365035
H2-D1	histocompatibility 2, D region [MGI:95912]	41%	39%	47%	41%	31%	40±6%	17	35400039	35404440
Il1rl1	interleukin 1 receptor-like 1 [MGI:98427]	43%	36%	33%	40%	32%	37±5%	1	40496621	40522241
Lmna	lamin A [MGI:96794]	35%	35%	45%	38%	33%	37±5%	3	88285071	88307229
Ltbp3	latent transforming growth factor beta binding protein 3 [MGI:1101355]	35%	43%	41%	37%	39%	39±3%	19	5740904	5758532
Msh2	mutS homolog 2 (*E. coli*) [MGI:101816]	23%	22%	21%	22%	22%	22±1%	17	88071880	88123052
Msh6	mutS homolog 6 (*E. coli*) [MGI:1343961]	3%	2%	3%	3%	2%	3±0%	17	88374424	88390403
Prkg2	protein kinase, cGMP-dependent, type II [MGI:108173]	48%	41%	48%	41%	42%	44±4%	5	99360287	99466098
Tnfrsf22	tumour necrosis factor receptor superfamily, member 22 [MGI:1930270]	35%	35%	37%	30%	37%	35±3%	7	150820717	150835557

A

	B6	D4	D8	F4	H14
Signal in mutant/ signal in *Blm*$^{-/-}$ cells	3.05	3.46	2.02	2.20	2.33
Adjusted P-value	2.77×10^{-20}	1.97×10^{-22}	$3.07\times^{-10}$	4.98×10^{-14}	2.10×10^{-14}

B

scl0012865.1_86-S probe sequence

GGCAGAGAAGCAGAAGCTCTTCCAGGCCGACAATGACCTCCCAGTACACT

1 23 26 50

C

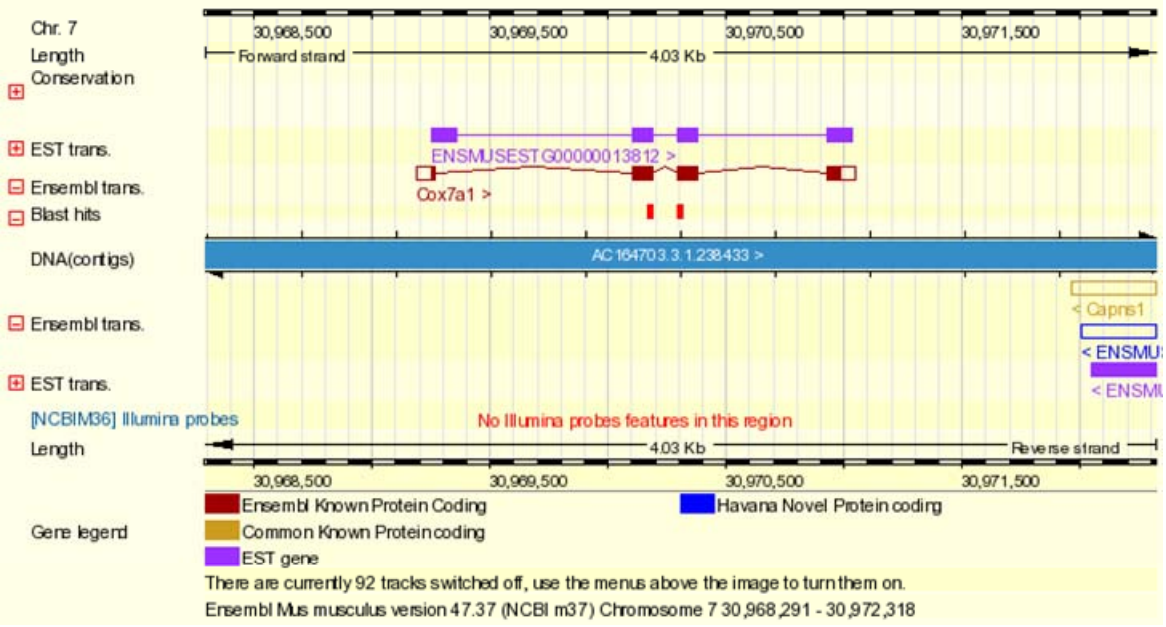


Figure 8-6 *Cox7a1* is up-regulated in all five *Msh2* and *Msh6* homozygously deleted mutants compared with the *Blm*-deficient NGG5-3 cell line

A. Comparison of the signals of probe scl0012865.1_86-S, the only probe for gene *Cox7a1*. The signal of this probe in any mutant is compared with its signal in the NGG5-3 *Blm*-deficient cell line. Fold changes are indicated. The adjusted P-value is used to indicate the possibility that this fold change is incorrect. **B**. Sequence of probe scl0012865.1_86-S. The red font will be explained in C. **C**. Probe scl0012865.1_86-S annotated in the mouse genome (NCBI m37). An Ensembl snapshot is shown here. The Illumina DAS source has not correctly annotated this probe in the genome, therefore Blast analysis was conducted. Two fragments of the probe sequence are 100% identical to the gene's sequence in exons 2 and 3. These two fragments are shown in red boxes. The first fragment contains nucleotide 1–26 of the probe. The second fragment contains nucleotide 23–50 of the probe. The red characters in B. indicate the overlapped nucleotides between two fragments.

8.2.1.5 *Msh6* homozygous deletion in mutant D1

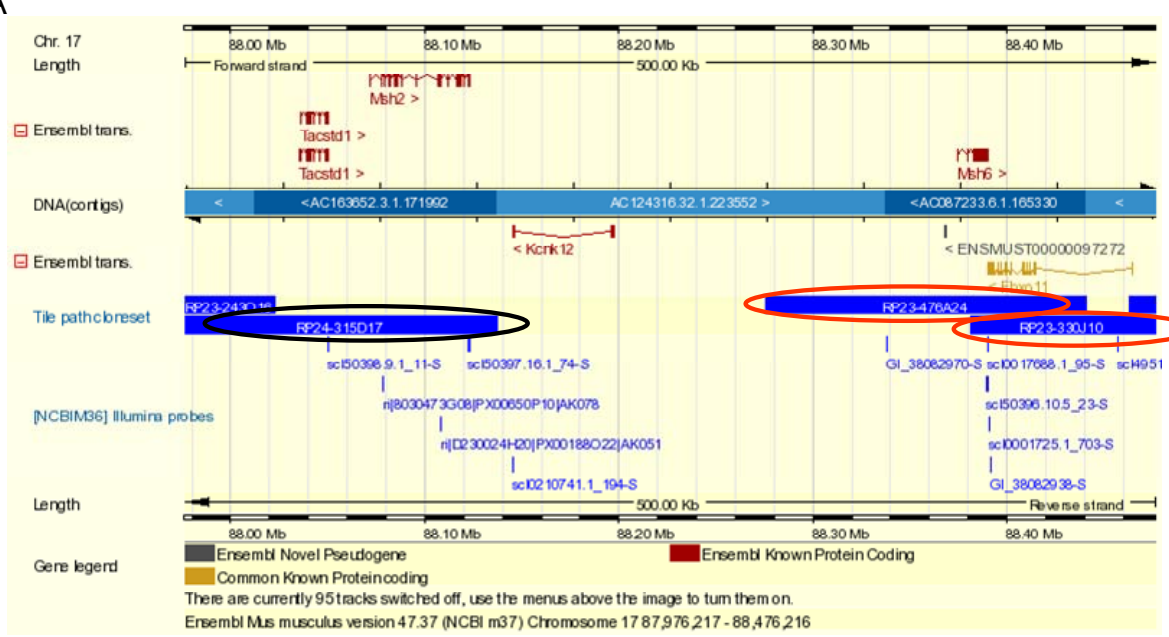
One mutant clone, D1, has lost both alleles of the *Msh6* gene. As *Msh2* is not affected in this clone, this clone was used to investigate downstream gene expression differences caused by *Msh6* loss and the differences compared with loss of both *Msh2* and *Msh6*. The D1 clone is strongly resistant to 6TG and has a 160 kb homozygous deletion from 87.8 Mb to 87.9 Mb on chromosome 17 with a Log_2(ratio) of -1.30 (Figure 8-7). By comparing probe signals on the expression array between the *Blm*-deficient cell line NGG5-3 and D1, a series of expression variations were identified. Most of the genes identified by comparing clones which had a deletion of both *Msh2* and *Msh6* (Table 8-6) were found to vary in this clone, too. The genes down-regulated by more than 50% in both D1 and the five *Msh2* and *Msh6* homozygous deletion mutants listed in Table 8-7.

Compared with the previous analysis (Table 8-6), three genes, *Bst2*, *Eif2s3y* and *Msh2,* were not altered in the D1 clone. The *Eif2s3y* gene is Y-linked, considering that the five mutants with *Msh2* and *Msh6* homozygous deletions had all lost the Y chromosome while the D1 mutant had not, it is reasonable that this gene is not altered. *Msh2* is not deleted in D1, thus it is also expected that this gene was found to be unaltered, too. The expression of *Bst2* (bone marrow stromal cell antigen 2) appears to be differentially affected by the activity of *Msh2*. Bone marrow stromal cells can regulate B-cell development through their surface molecules (Ishikawa *et al.* 1995).

The analysis of genes with increased expression in the D1 clone also identified the cytochrome c oxidase, subunit VIIa 1 (*Cox7a1)* gene which doubled (2.08 times) its expression signal in the D1 mutant compared with the *Blm*-deficient cell line NGG5-3. This confirms that *Cox7a1* is transcriptionally upregulated in MMR-deficient ES cells. Similarly, the expression signal of this gene in the *Dnmt1* knockout cells (ww56+*Hprt*) is 1.48 times more than that in the AB2.2 wild type cells.

When comparing the down-regulated genes in the *Msh6* mutant D1 and the five mutant clones with the *Msh2* and *Msh6* homozygous deletions with the down-regulated genes in the *Dnmt1* knockout control cell line ww56+Hprt, no gene appears in both comparisons. This indicates that the gene expression changes in *Dnmt1* and *Msh2/Msh6* mutants are different.

A



B

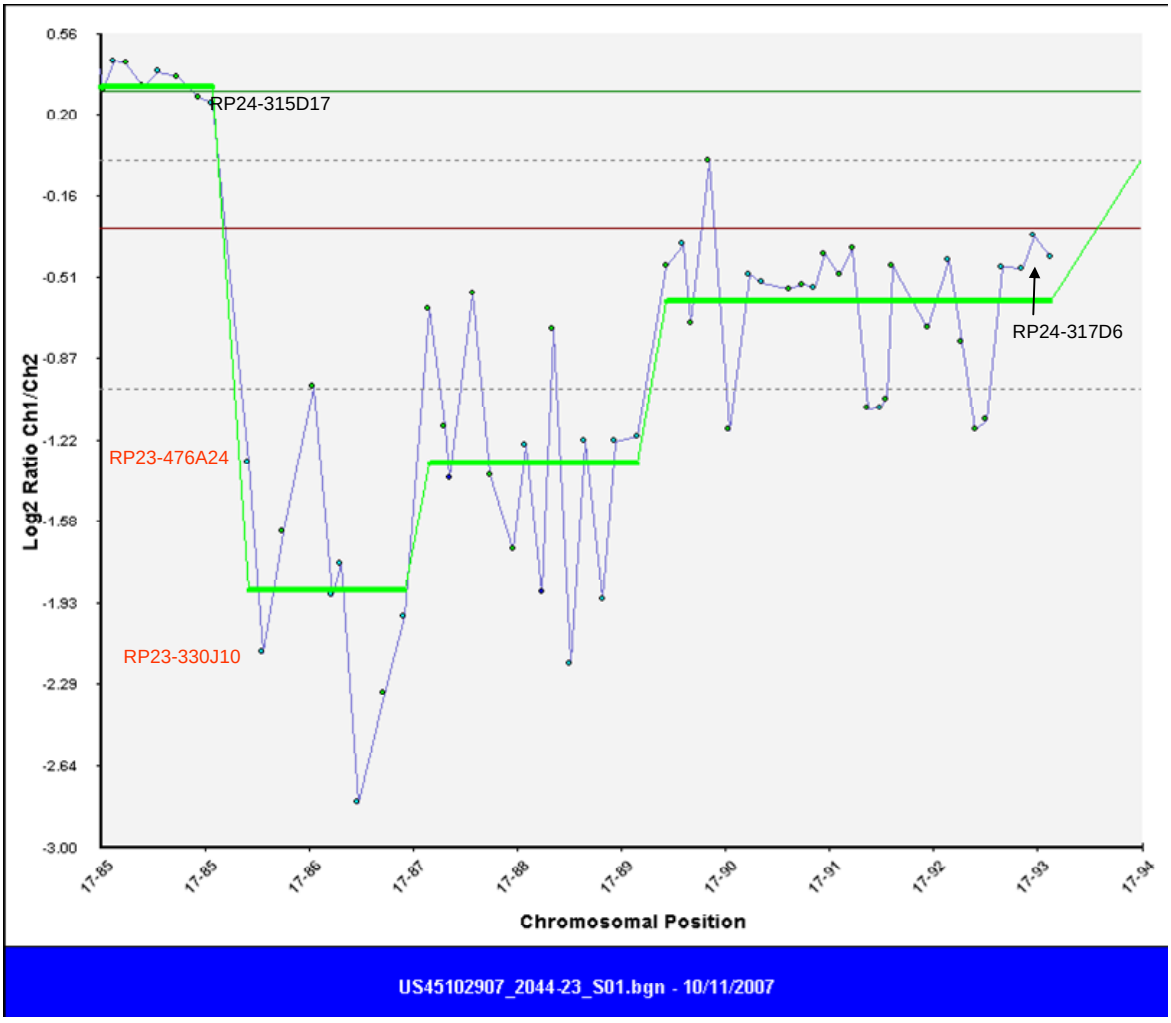


Figure 8-7 Analysis of the mutant D1, a clone with an *Msh6* homozygous deletion

Figure 8-7 Analysis of the mutant D1, a clone with *Msh6* homozygous deletion

A. An Ensembl snapshot of a 500 kb region showing the *Msh2* and *Msh6* genes. Ensembl annotated transcripts are shown in dark red. Tile path BACs for array CGH are shown as blue boxes. The black circled BAC, RP24-315D17, covers the *Msh2* gene. Two red circled BACs, RP23-476A24 and RP23-330J10, cover the *Msh6* gene. There is a gap in the tile path between RP24-315D17 and RP23-476A24. Gene-specific probes on the Illumina® array are shown as small blue bars. **B**. Zoomed-in array CGH profile of distal chromosome 17. BAC RP24-315D17 does not exhibit any copy number change compared with the *Blm*-deficient cell line NGG5-3. However, BAC RP23-476A24 and RP23-330J10 (red font) are within a homozygous deletion. BAC RP24-316D6 is the last tile path array BAC on chromosome 17. The Log_2(ratio) does not confirm the profile expected for a homozygous deletion. This clone is likely to have a complex deletion from BAC RP23-476A24 to the end of chromosome 17. One interpretation of this pattern is that one allele has deleted 160 kb, including *Msh6,* while the other has suffered a much larger deletion. Thus, only the 160 kb region shows a homozygous deletion.

Table 8-7 Genes down-regulated in both D1 (*Msh6*$^{-/-}$ deletion) and the five *Msh2* and *Msh6* deficient mutants compared to *Blm*-deficient ES cells

MGI symbol	Description	% *Blm*$^{-/-}$ signal						Average % *Blm*-/- signal ± SD	Chromosome Name
		B6	D1	D4	D8	F4	H14		
2310016C16Rik	RIKEN cDNA 2310016C16 gene [MGI:1916840]	44%	27%	28%	31%	37%	31%	33±6%	13
Ank	progressive ankylosis [MGI:3045421]	41%	37%	35%	42%	38%	39%	39±3%	15
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	47%	41%	38%	40%	46%	35%	41±5%	4
Col6a1	procollagen, type VI, alpha 1 [MGI:88459]	50%	39%	42%	44%	45%	29%	42±7%	10
Cpe	carboxypeptidase E [MGI:101932]	50%	36%	40%	40%	46%	38%	42±5%	8
Dcn	decorin [MGI:94872]	33%	20%	22%	27%	25%	28%	26±5%	10
Ecm1	extracellular matrix protein 1 [MGI:103060]	39%	45%	47%	48%	44%	43%	44±3%	3
H2-D1	histocompatibility 2, D region [MGI:95912]	41%	40%	39%	47%	41%	31%	40±5%	17
Il1rl1	interleukin 1 receptor-like 1 [MGI:98427]	43%	32%	36%	33%	40%	32%	36±5%	1
Lmna	Lamin A [MGI:96794]	35%	36%	35%	45%	38%	33%	37±4%	3
Ltbp3	Latent transforming growth factor beta binding protein 3 [MGI:1101355]	35%	40%	43%	41%	37%	39%	39±3%	19
Msh6	mutS homolog 6 (*E. coli*) [MGI:1343961]	3%	2%	2%	3%	3%	2%	3±0%	17
Prkg2	protein kinase, cGMP-dependent, type II [MGI:108173]	48%	43%	41%	48%	41%	42%	44±3%	5
Tnfrsf22	Tumor necrosis factor receptor superfamily, member 22 [MGI:1930270]	35%	34%	35%	37%	30%	37%	34±3%	7

The above 14 genes are down-regulated more than 50% in the mutants of B6, D4, D8, F4, H14 and D1, comparing to the expression levels of all genes in the *Blm*-deficient control ES cell line NGG5-3. The individual and the average expression levels of the mutants compared to the *Blm*-deficient cells are shown with standard deviation (SD).

8.2.1.6 Expression status of *Msh2* and *Msh6* in heterozygous deletions

There were nine mutants (C1, C3, C5, D6, F9, F16, G8, G9 and H13) which had heterozygous deletions covering both *Msh2* and *Msh6* genes (see array CGH analysis). Six of them (C1, C5, F16, G8, G9 and H13) are strongly resistant to 6TG. Three (C3, D6 and F9) are weakly resistant to 6TG. F9 and F16 are clonal thus expression array analysis was performed on one of them, F9. As heterozygous mutation of *Msh2* does not lead to resistance to 6TG (Abuin *et al.* 2000; de Wind *et al.* 1995) and one copy of *Msh2* and *Msh6* should be able to produce functional proteins, either *Msh2* or *Msh6* are expected to have a mutation in the other allele, which may result in a loss of gene expression caused by changes in the reading frame and generation of premature termination codons (PTCs). Using these PTCs as signals, the nonsense-mediated decay (NMD) mechanism can selectively degrade mutant mRNA.

Expression array analysis identified a number of mutants which had the same magnitude of signal reduction for *Msh2* or *Msh6* as in the homozygous deletions (Table 8-8). The *Msh2* probe scl50397.16.1_74-S and the *Msh6* probe scl0017688.1_95-S are reliable probes as these two probes show 2.5– 25% expression signals of that in the *Blm*-deficient control cells, while the other probes of these two genes exhibit 65–75% expression signals or inconsistent signals (probe ri|D230024H20|PX00188O22|AK051) in the homozygous deletions compared with the *Blm*-deficient control cells (Figure 8-4 and Figure 8-5). The expression level of the *Msh2* probe scl50397.16.1_74-S in the mutant C5 and H13 is 25–26% of that in the *Blm*-deficient control cells. This expression level is similar to those in the homozygous deletions (21–22%). Additionally, the expression level of the *Msh6* probe scl0017688.1_95-S in the mutant G8 is 2% of that in the *Blm*-deficient control cells, the same signal level of that in the homozygously deleted mutants. These data expain why the mutant clones C5, H13 and G8 are strongly resistant to 6TG. These results indicate that these clones carry mutations in the other allele of *Msh2* or *Msh6*. However, one clone (G9) shows a reduced magnitude of reduction of the *Msh6* signal (nine fold) compared to that seen in homozygous mutations and no reduction in the *Msh2* signal. This clone is believed to be an *Msh6* mutant, too. This explains the strong resistance to 6TG of the clone G9. In a summary, six homozygous mutants for the *Msh2* and *Msh6* genes have lost the expression signals of the both genes. Two heterozygous mutants (C5 and H13) lost the transcripts of *Msh2* and two heterozygous mutants (G8 and G9) lost or significantly reduced the expression of *Msh6*. These four mutants may contain compound mutations of deletion/small mutation in either *Msh2* or *Msh6*, thus are strongly resistant to 6TG. While the cause of the strong 6TG resistance of the mutant C1 remains unclear. But the weak 6TG resistance of

Table 8-8 Expression signal decrease of *Msh2* and *Msh6* probes in homozygously and heterozygously deleted mutants

A

	Clone	6TG^R	*Msh2* probe scl50397.16.1_74-S	
			Expression fold change (NGG5-3/mutant)	% *Blm*$^{-/-}$ signal
Homozygous deletion	B6	Strong	4.32	23%
	D4	Strong	4.63	22%
	D8	Strong	4.79	21%
	F4	Strong	4.47	22%
	H14	Strong	4.56	22%
Heterozygous deletion	C1	Strong	1.97	51%
	C3	Weak	2.01	50%
	C5	Strong	3.81	26%
	D6	Weak	2.25	44%
	F9	Weak	NA	NA
	G8	Strong	1.78	56%
	G9	Strong	NCD	NCD
	H13	Strong	4.06	25%

B

	Clone	6TG^R	*Msh6* probe scl0017688.1_95-S	
			Expression fold change (NGG5-3/mutant)	% *Blm*$^{-/-}$ signal
Homozygous deletion	B6	Strong	30.7	3%
	D1	Strong	42.8	2%
	D4	Strong	43.4	2%
	D8	Strong	38.9	3%
	F4	Strong	36.3	3%
	H14	Strong	46.9	2%
Heterozygous deletion	C1	Strong	NCD	NCD
	C3	Weak	NCD	NCD
	C5	Strong	1.67	60%
	D6	Weak	1.38	72%
	F9	Weak	NA	NA
	G8	Strong	49.5	2%
	G9	Strong	9.32	11%
	H13	Strong	1.5	67%

A. Signal decrease of an *Msh2* probe in mutants with *Msh2* and *Msh6* homozygous and heterozygous deletions. The phenotype of 6TG resistance is shown. NA: data not available. NCD: no convincing data (comparison P-value is high). **B**. Signal decrease of an *Msh6* probe in mutants with *Msh2* and *Msh6* heterozygous deletions.

the mutant clones C3, D6 and F9 may result from the hypomorphic mutation of the *Msh2* and *Msh6* genes.

By comparing all the mutants, five genes were found to be down-regulated to 20–50% in all of the *Msh2* and/or *Msh6* homozygous and compound mutants (Table 8-9). These genes are *Col16a1* (procollagen, type XVI, alpha 1), *Dcn* (decorin), *Lmna* (Lamin A), *Ltbp3* (Latent transforming growth factor beta binding protein 3) and *Tnfrsf22* (Tumor necrosis factor receptor superfamily, member 22). A common feature of these genes is that their products may locate on or close to cell membrane and nuclear membrane (Lamin A). The down-regulation of these genes may be secondary effects caused by MMR defects.

It is known that the cytochrome c oxidase, subunit VIIa 1 (*Cox7a1)* gene probe scl0012865.1_86-S was detected which showed 2.0–3.5 times increased expression in the six *Msh2/Msh6* homozygous mutants (B6, D1, D4, D8, F4 and H14). When comparing the compound heterozygous mutants (C5, H13, G8 and G9) which lost the function of *Msh2* or *Msh6* to detect upregulated genes, the *Cox7a1* gene is one of the ten genes up-regulated in the C5, G8 and G9 mutants. The expression levels of this gene in C5, G8 and G9 are 2.3–3.7 times of that in the *Blm*-deficient control cells. While the probe performance of *Cox7a1* was not so consistent in the mutant H13 because the comparison P-value of this probe is higher than the lowest five thousand probes; therefore, this probe was not included in these comparisons. But it is possible that *Cox7a1* is also over expressed in the H13 clone. In a summary, the *Cox7a1* gene showed 2.0–3.7 times increased expression in the nine *Msh2/Msh6* homozygous or compound mutants (B6, C5, D1, D4, D8, F4, G8, G9 and H14).

Table 8-9 Genes down-regulated in ten *Msh2/Msh6* deficient mutants compared to *Blm*-deficient ES cells

MGI symbol	Description	% *Blm*$^{-/-}$ signal										Average % *Blm*-/- signal ± SD	Chromosome Name
		B6	C5	D1	D4	D8	F4	G8	G9	H13	H14		
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	47%	47%	41%	38%	40%	46%	30%	38%	46%	35%	41±6%	4
Dcn	decorin [MGI:94872]	33%	35%	20%	22%	27%	25%	21%	21%	43%	28%	27±7%	10
Lmna	Lamin A [MGI:96794]	35%	37%	36%	35%	45%	38%	26%	27%	33%	33%	34±5%	3
Ltbp3	Latent transforming growth factor beta binding protein 3 [MGI:1101355]	35%	46%	40%	43%	41%	37%	31%	38%	46%	39%	40±5%	19
Tnfrsf22	Tumor necrosis factor receptor superfamily, member 22 [MGI:1930270]	35%	36%	34%	35%	37%	30%	24%	32%	49%	37%	35±6%	7

The above five genes are down-regulated to 20–50% in the mutants of B6, C5, D1, D4, D8, F4, G8, G9, H13, and H14, comparing to the expression levels in the *Blm*-deficient control ES cell line NGG5-3. The individual and the average expression levels of the mutants compared to the *Blm*-deficient cells are shown with standard deviation (SD).

8.2.2 Combined analysis of array CGH and expression array results

8.2.2.1 Positive relationship between copy number and expression variation

To test the usefulness of a combined analysis of array CGH and expression array data, a small-scale analysis was conducted to examine the relationship between a given gene's transcription level and its copy number. The samples in this small-scale experiment included mutant clones B2 (weak $6TG^R$), E1 (strong $6TG^R$), E2 (weak $6TG^R$), F1 (strong $6TG^R$) and G4 (weak $6TG^R$). These clones were selected because these mutant clones had all deleted one allele of a 12 Mb genomic region on chromosome 14 (107.9 Mb–119.6 Mb, Figure 7-25). It was expected that genes homozygously deleted in the mutants should show reduced expression while amplified genes should express at higher levels. The array signals of genes in the amplified and deleted regions in the five mutants were compared with array signals of *Blm*-deficient cells. The Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) data from the expression array is used to indicate the expression of a given gene compared with the *Blm*-deficient control cell line NGG5-3 and the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) data from the array CGH is used to indicate the copy number change of a given gene. By plotting these two values in a figure, the relationship between copy number and expression level can be discovered.

From this analysis (Figure 8-8), it was found that genes in amplified regions express more mRNA than in the control cell line; and most of the genes express less mRNA than the control cell line does when the region is deleted. Thus, there is a positive relationship between copy number and expression level in most of the cases. It is noticeable that one gene is homozygously deleted in four out of five mutant clones (except E1) with the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) being less than minus one. This gene is *Eif2s3y*, eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked. This kind of gene could be the best MMR candidate resulting from the screen. However, this gene is on the Y chromosome and the Y chromosome is lost in all four of these clones. Clone E1 has an intact Y chromosome thus does not show this phenomenon.

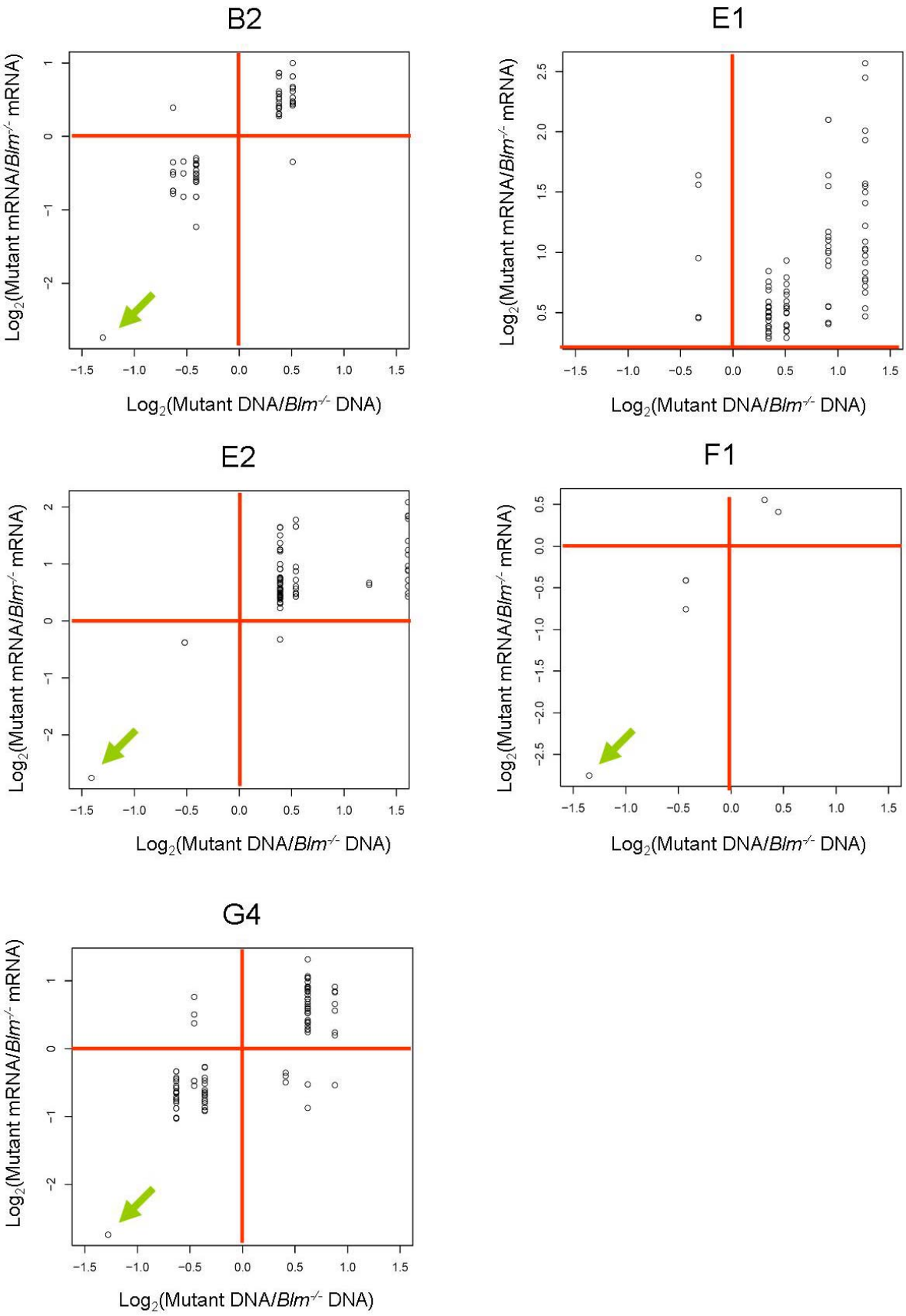


Figure 8-8 **Positive relationship between copy number and expression level**

Figure 8-8 Positive relationship between copy number and expression level

In mutants B2, E1, E2, F1 and G4, data of expression change of amplified and deleted genes is plotted against data of their copy number changes. Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) is used to indicate the expression variations of mutant genes compared with the expression of their counterparts in the *Blm*-deficient control cell line NGG5-3. The Y axis measures Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA). A given mutant gene is overexpressed when the Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) is greater than zero; and it is down-regulated when the Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) is less than zero. The X axis measures Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA). These data are from the array CGH analysis. Similarly, a given gene is amplified when the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) is greater than zero; and it is deleted when the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) is less than zero. The red lines indicate zero on both the X and Y axes. Most of the genes fall into either the (Plus, Plus) region or the (Minus, Minus) region, indicating that in general, when a gene gains more copies in the genome, it expresses more; when a gene loses one or both copies of its sequence, its transcription is also decreased. The green arrows indicate that in four out of five mutant clones (except E1), gene *Eif2s3y* is deleted and its expression is shut down.

8.2.2.2 Genes affected by MMR defects

Using a similar strategy to the previous analysis section, the expression level of genes in deleted regions can be examined within thirty-four $6TG^R$ mutant clones, which were analysed using array CGH. It was assumed that if a gene was homozygously deleted and its transcription was lost in different mutant clones, it would be likely that the deficiency of this gene would be associated with 6TG resistance and MMR. However, a loss-of-function mutation may be caused by a large heterozygous deletion combined with a small mutation at the relevant locus. This type of combination of mutations cannot be shown as a homozygous deletion. Thus, in this analysis, any gene was selected to examine the decrease of its expression if they were homozygously deleted in one mutant clone. The transcription levels of these genes in the mutants were separately compared with the NGG5-3 *Blm*-deficient cells.

By combining data from both array CGH analysis and expression array analysis, 34 genes were found to be homozygously deleted in at least one mutant and their transcripts can be observed in the *Blm*-deficient ES cells, but the expression level of these genes in the $6TG^R$ mutants was less than 50% of that in the *Blm*-deficient cells (Table 8-10). The genes in this list may be indirectly affected by MMR defects or 6TG metabolism pathway. The frequency of homozygous deletion of genes and the number of $6TG^R$ mutants in which each gene is down-regulated are shown in the Table 8-10. The *Msh2* and *Msh6* genes can be found in Table 8-10, both being down-regulated in nine $6TG^R$ mutants. Some genes on the list are on chromosome 17, which could be associated with the *Msh2* and *Msh6* homozygous deletions. Examples are the *Cox7a2l*, *Fbxo11*, *Ppm1b* and *Socs5* genes, which are down-regulated in ten to twenty-one $6TG^R$ mutants and are within a short distance of the *Msh2* and *Msh6* genes. All genes down-regulated on chromosome 17 are within a 32 Mb region, ranging from 56.4 Mb (*Uhrf1*) to 88.4 Mb (*Fbxo11*). These genes are down-regulated probably because they were deleted together with the *Msh2* and *Msh6* genes. However, this does not rule out the possibility that these genes and other genes on chromosome 17 may be involved in MMR, such as genes *Bcat1* (Ch6, reduced expression in 10 clones), *Eps8* (Ch6, reduced expression in 6 clones), *Eif5* (Ch12, reduced expression in 7 clones), *Ank* (Ch15, reduced expression in 22 clones) and *Dap* (Ch15, reduced expression in 8 clones) are not on chromosome 17 and were down-regulated in between 6 (*Eps8*) and 22 (*Ank*) mutant clones. This provides some evidence that these genes affected by DNA mismatch repair defects. The *Zfp459* (Ch13, reduced expression in 8 clones) gene seems to be homozygously deleted for seven times. However, this gene is located on a BAC,

which should belong to chromosome Y instead of chromosome 13. Its decreased expression may be associated with loss of chromosome Y.

In addition, over-expression of the human homologue *EPS8* was observed in pancreatic cancer, that can facilitate colon cancer cell growth (Maa *et al.* 2007; Welsch *et al.* 2007). The EPS8 protein may function through the expression of the focal adhesion kinase (*Fak*) gene (Maa *et al.* 2007). Among these six genes, only *Dap* has been shown to be associated with MMR in the literature. The promoter region of death-associated protein (DAP) kinase gene (*DAPK-1*) was observed to be hypermethylated in early-onset sporadic gastric carcinoma and B-cell lymphoma cell lines (Katzenellenbogen *et al.* 1999; Kim *et al.* 2005; Kissil *et al.* 1997). The expressional down-regulation of *DAPK-1* due to hypermethylation indicates that this gene might play a role in tumour suppression.

Because the NGG5-3 *Blm*-deficient cells have been modified by multiple steps of gene targeting, and *Blm* deficiency causes genome instability to the cells, these cells may contain unknown mutations. Therefore, the above analysis was also conducted on wild type cells (both AB1 and AB2.2 cell lines, two replicates each). The genes, which were homozygously deleted in at least one $6TG^R$ mutant, were compared at a transcriptional level between any single mutant and the wild type cells. The number of mutants in which gene expression is down-regulated compared to the wild type cells is shown. These genes (42) are listed in Table 8-11. Several genes were identified from this comparison which had not been identified in the comparison with the *Blm*-deficient line NGG5-3. For example, the *Amn* and *1810015C04Rik* genes show reduced transcription in 26 (*Amn*) or 21 (*1810015C04Rik*) of thirty-one $6TG^R$ mutants analysed. The reason for such a high frequency and difference between the *Blm*-deficient cells and the wild type cells is due to their lower expression in the *Blm*-deficient reference cells and the *Blm*-deficient mutant cells. Then expression differences may be spontaneously introduced by *Blm* deficiency during multiple targeting procedures.

Twenty-six of the fourty-two genes in Table 8-11 overlapped with the genes in Table 8-10. The 26 overlapping genes are shown in Table 8-12, in which genes were homozygously deleted in at least one $6TG^R$ mutant and the gene expression in some mutants was reduced to less than 50% of that in both the *Blm*-deficient cells and the wild type cells. Again, genes *Msh2*, *Msh6* and other genes on chromosome 17 are in Table 8-12. Five genes on chromosome 15 and sixteen genes on chromosome 17 are found in these 26 genes and are illustrated in Figure 8-9. *Bcat1, Eps8* and *Dap* are not in Table 8-11 and Table 8-12. An

explanation may be that the expression level of these three genes has a large range of variation. The *Eif5* and *Ank* genes are in Table 8-12 with a relatively higher frequency of being down-regulated in the $6TG^R$ mutants. This indicates that they are more likely to be involved in or affected by MMR deficiency.

Table 8-10 Genes which are deleted in one mutant and whose transcripts are down-regulated in mutants compared with the *Blm*-deficient cells

MGI symbol	Description	Chromosome Name	Gene Start (bp)	Gene End (bp)	Frequency in deletion	Frequency of down-regulation
Ank	progressive ankylosis [MGI:3045421]	15	27396432	27524660	1	22
Socs5	suppressor of cytokine signaling 5 [MGI:2385459]	17	87507019	87536923	5	21
Fbxo11	F-box protein 11 [MGI:2147134]	17	88390199	88464625	12	10
Cox7a2l	cytochrome c oxidase subunit VIIa polypeptide 2-like [MGI:106015]	17	83901270	83913670	2	10
Ppm1b	protein phosphatase 1B, magnesium dependent, beta isoform [MGI:101841]	17	85357341	85423074	2	10
Bcat1	branched chain aminotransferase 1, cytosolic [MGI:104861]	6	144947678	145024704	1	10
Msh6	mutS homolog 6 (E. coli) [MGI:1343961]	17	88374424	88390403	12	9
Msh2	mutS homolog 2 (E. coli) [MGI:101816]	17	88071880	88123052	6	9
Zfp459	zinc finger protein 459 [MGI:3040701]	13	67506658	67514893	7	8
Dap	death-associated protein [MGI:1918190]	15	31154147	31204095	1	8
Eif5	eukaryotic translation initiation factor 5 [MGI:95309]	12	112776312	112784822	1	7
Fez2	fasciculation and elongation protein zeta 2 (zygin II) [MGI:2675856]	17	78777225	78817471	2	6
Eps8	epidermal growth factor receptor pathway substrate 8 [MGI:104684]	6	137425766	137597806	1	6
Ppp4r1	protein phosphatase 4, regulatory subunit 1 [MGI:1917601]	17	66132738	66191187	2	4
Vapa	vesicle-associated membrane protein, associated protein A [MGI:1353561]	17	65929396	65962895	1	4
Bmp4	bone morphogenetic protein 4 [MGI:88180]	14	47003195	47010274	1	3
1110012J17Rik	RIKEN cDNA 1110012J17 gene [MGI:1915867]	17	66686326	66799087	1	3
Gemin6	gem (nuclear organelle) associated protein 6 [MGI:1914492]	17	80623791	80627810	2	2
Map4k3	mitogen-activated protein kinase kinase kinase kinase 3 [MGI:2154405]	17	80979852	81127433	2	2
Rab31	RAB31, member RAS oncogene family [MGI:1914603]	17	66001070	66122092	2	2
Uhrf1	ubiquitin-like, containing PHD and RING finger domains, 1 [MGI:1338889]	17	56442823	56462909	2	2
Ldhb	lactate dehydrogenase B [MGI:96763]	6	142438772	142456477	1	2
Hsp90aa1	heat shock protein 90kDa alpha (cytosolic), class A member 1 [MGI:96250]	12	111929238	111934605	1	2
Grhl2	grainyhead-like 2 (*Drosophila*) [MGI:2182543]	15	37163002	37293317	1	2
Ttc27	tetratricopeptide repeat domain 27 [MGI:1921446]	17	75117090	75262910	1	2
Clpp	caseinolytic peptidase, ATP-dependent, proteolytic subunit homolog (E. coli) [MGI:1858213]	17	57129725	57135610	3	1

Arl6ip2	ADP-ribosylation factor-like 6 interacting protein 2 [MGI:1929492]	17	80247732	80295463	2	1
Epb4.1l3	erythrocyte protein band 4.1-like 3 [MGI:103008]	17	69506150	69639325	2	1
Mta3	metastasis associated 3 [MGI:2151172]	17	84105503	84223055	2	1
Mylc2b	myosin light chain, regulatory B [MGI:107494]	17	71323282	71339897	2	1
Entpd4	ectonucleoside triphosphate diphosphohydrolase 4 [MGI:1914714]	14	69955231	69985327	1	1
Atp6v1c1	ATPase, H+ transporting, lysosomal V1 subunit C1 [MGI:1913585]	15	38591699	38622197	1	1
Mtdh	Metadherin [MGI:1914404]	15	34012474	34072140	1	1
Pabpc1	poly A binding protein, cytoplasmic 1 [MGI:1349722]	15	36525416	36538728	1	1

Table 8-11 Genes which are deleted in one mutant and whose transcripts are down-regulated in mutants compared with wild type cells (AB1 and AB2.2)

MGI symbol	Description	Chromosome Name	Gene Start (bp)	Gene End (bp)	Frequency of deletion vs *Blm*$^{-/-}$ cells	Frequency of down-regulation vs wild type cells
Amn	amnionless [MGI:1934943]	12	112509322	112514637	0	26
1810015C04Rik	RIKEN cDNA 1810015C04 gene [MGI:1913520]	15	25773019	25903442	0	21
Fez2	fasciculation and elongation protein zeta 2 (zygin II) [MGI:2675856]	17	78777225	78817471	2	11
Msh2	mutS homolog 2 (E. coli) [MGI:101816]	17	88071880	88123052	6	11
Msh6	mutS homolog 6 (E. coli) [MGI:1343961]	17	88374424	88390403	12	10
Ppp4r1	protein phosphatase 4, regulatory subunit 1 [MGI:1917601]	17	66132738	66191187	2	10
Grhl2	grainyhead-like 2 (*Drosophila*) [MGI:2182543]	15	37163002	37293317	1	9
Ppm1b	protein phosphatase 1B, magnesium dependent, beta isoform [MGI:101841]	17	85357341	85423074	2	8
Vapa	vesicle-associated membrane protein, associated protein A [MGI:1353561]	17	65929396	65962895	1	8
Zfp459	zinc finger protein 459 [MGI:3040701]	13	67506658	67514893	7	7
Fbxo11	F-box protein 11 [MGI:2147134]	17	88390199	88464625	12	7
Pabpc1	poly A binding protein, cytoplasmic 1 [MGI:1349722]	15	36525416	36538728	1	6
Crip2	cysteine rich protein 2 [MGI:1915587]	12	114378742	114383715	0	5
Rcor1	REST corepressor 1 [MGI:106340]	12	112277825	112351595	0	5
Wars	tryptophanyl-tRNA synthetase [MGI:104630]	12	110098244	110132355	0	5
1110012J17Rik	RIKEN cDNA 1110012J17 gene [MGI:1915867]	17	66686326	66799087	1	5
Eif5	eukaryotic translation initiation factor 5 [MGI:95309]	12	112776312	112784822	1	4
Arl6ip2	ADP-ribosylation factor-like 6 interacting protein 2 [MGI:1929492]	17	80247732	80295463	2	4
Cox7a2l	cytochrome c oxidase subunit VIIa polypeptide 2-like [MGI:106015]	17	83901270	83913670	2	4
1700019H03Rik	RIKEN cDNA 1700019H03 gene [MGI:1919107]	2	180459968	180476985	0	3
AW555464	expressed sequence AW555464 [MGI:2145043]	12	113960445	113984802	0	3
Hsp90aa1	heat shock protein 90kDa alpha (cytosolic), class A member 1 [MGI:96250]	12	111929238	111934605	1	3
Ppp2r5c	protein phosphatase 2, regulatory subunit B (B56), gamma isoform [MGI:1349475]	12	111685518	111821271	0	3
Eml4	echinoderm microtubule associated protein like 4 [MGI:1926048]	17	83750322	83879697	0	3
Mta3	metastasis associated 3 [MGI:2151172]	17	84105503	84223055	2	3
Mylc2b	myosin light chain, regulatory B [MGI:107494]	17	71323282	71339897	2	3
Rfx2	Regulatory factor X, 2 (influences HLA class II expression) [MGI:106583]	17	56915323	56970436	0	3
Zfyve21	zinc finger, FYVE domain containing 21 [MGI:1915770]	12	113052381	113066596	0	2
Ank	progressive ankylosis [MGI:3045421]	15	27396432	27524660	1	2
M6prbp1	mannose-6-phosphate receptor binding protein 1 [MGI:1914155]	17	56416899	56429934	0	2

Rab31	RAB31, member RAS oncogene family [MGI:1914603]	17	66001070	66122092	2	2
Uhrf1	ubiquitin-like, containing PHD and RING finger domains, 1 [MGI:1338889]	17	56442823	56462909	2	2
Ythdf1	YTH domain family 1 [MGI:1917431]	2	180639084	180655641	0	1
Ldhb	lactate dehydrogenase B [MGI:96763]	6	142438772	142456477	1	1
Bmpr1a	bone morphogenetic protein receptor, type 1A [MGI:1338938]	14	35224255	35315732	0	1
Entpd4	ectonucleoside triphosphate diphosphohydrolase 4 [MGI:1914714]	14	69955231	69985327	1	1
Atp6v1c1	ATPase, H+ transporting, lysosomal V1 subunit C1 [MGI:1913585]	15	38591699	38622197	1	1
Hrsp12	heat-responsive protein 12 [MGI:1095401]	15	34413776	34424935	0	1
Mtdh	Metadherin [MGI:1914404]	15	34012474	34072140	1	1
Epb4.1l3	erythrocyte protein band 4.1-like 3 [MGI:103008]	17	69506150	69639325	2	1
Map4k3	mitogen-activated protein kinase kinase kinase kinase 3 [MGI:2154405]	17	80979852	81127433	2	1
Sfrs7	splicing factor, arginine/serine-rich 7 [MGI:1926232]	17	80599426	80606643	0	1

Table 8-12 Genes which are deleted in one mutant and whose transcripts are down-regulated in mutants compared with wild type cells and the *Blm*-deficient cells

MGI symbol	Description	Chromosome Name	Gene Start (bp)	Gene End (bp)	Frequency of deletions vs *Blm*$^{-/-}$ cells	Frequency of down-regulation vs wild type cells	Frequency of down-regulation vs Blm$^{-/-}$ cells
Ank	progressive ankylosis [MGI:3045421]	15	27396432	27524660	1	2	22
Cox7a2l	cytochrome c oxidase subunit VIIa polypeptide 2-like [MGI:106015]	17	83901270	83913670	2	4	10
Fbxo11	F-box protein 11 [MGI:2147134]	17	88390199	88464625	12	7	10
Ppm1b	protein phosphatise 1B, magnesium dependent, beta isoform [MGI:101841]	17	85357341	85423074	2	8	10
Msh2	mutS homolog 2 (E. coli) [MGI:101816]	17	88071880	88123052	6	11	9
Msh6	mutS homolog 6 (E. coli) [MGI:1343961]	17	88374424	88390403	12	10	9
Zfp459	zinc finger protein 459 [MGI:3040701]	13	67506658	67514893	7	7	8
Eif5	Eukaryotic translation initiation factor 5 [MGI:95309]	12	112776312	112784822	1	4	7
Fez2	fasciculation and elongation protein zeta 2 (zygin II) [MGI:2675856]	17	78777225	78817471	2	11	6
Ppp4r1	protein phosphatise 4, regulatory subunit 1 [MGI:1917601]	17	66132738	66191187	2	10	4
Vapa	vesicle-associated membrane protein, associated protein A [MGI:1353561]	17	65929396	65962895	1	8	4
1110012J17Rik	RIKEN cDNA 1110012J17 gene [MGI:1915867]	17	66686326	66799087	1	5	3
Ldhb	lactate dehydrogenase B [MGI:96763]	6	142438772	142456477	1	1	2
Hsp90aa1	heat shock protein 90kDa alpha (cytosolic), class A member 1 [MGI:96250]	12	111929238	111934605	1	3	2
Grhl2	grainyhead-like 2 (*Drosophila*) [MGI:2182543]	15	37163002	37293317	1	9	2
Map4k3	mitogen-activated protein kinase kinase kinase kinase 3 [MGI:2154405]	17	80979852	81127433	2	1	2
Rab31	RAB31, member RAS oncogene family [MGI:1914603]	17	66001070	66122092	2	2	2
Uhrf1	ubiquitin-like, containing PHD and RING finger domains, 1 [MGI:1338889]	17	56442823	56462909	2	2	2
Entpd4	ectonucleoside triphosphate diphosphohydrolase 4 [MGI:1914714]	14	69955231	69985327	1	1	1
Atp6v1c1	ATPase, H+ transporting, lysosomal V1 subunit C1 [MGI:1913585]	15	38591699	38622197	1	1	1

Mtdh	Metadherin [MGI:1914404]	15	34012474	34072140	1	1	1
Pabpc1	poly A binding protein, cytoplasmic 1 [MGI:1349722]	15	36525416	36538728	1	6	1
Arl6ip2	ADP-ribosylation factor-like 6 interacting protein 2 [MGI:1929492]	17	80247732	80295463	2	4	1
Epb4.1l3	erythrocyte protein band 4.1-like 3 [MGI:103008]	17	69506150	69639325	2	1	1
Mta3	metastasis associated 3 [MGI:2151172]	17	84105503	84223055	2	3	1
Mylc2b	myosin light chain, regulatory B [MGI:107494]	17	71323282	71339897	2	3	1

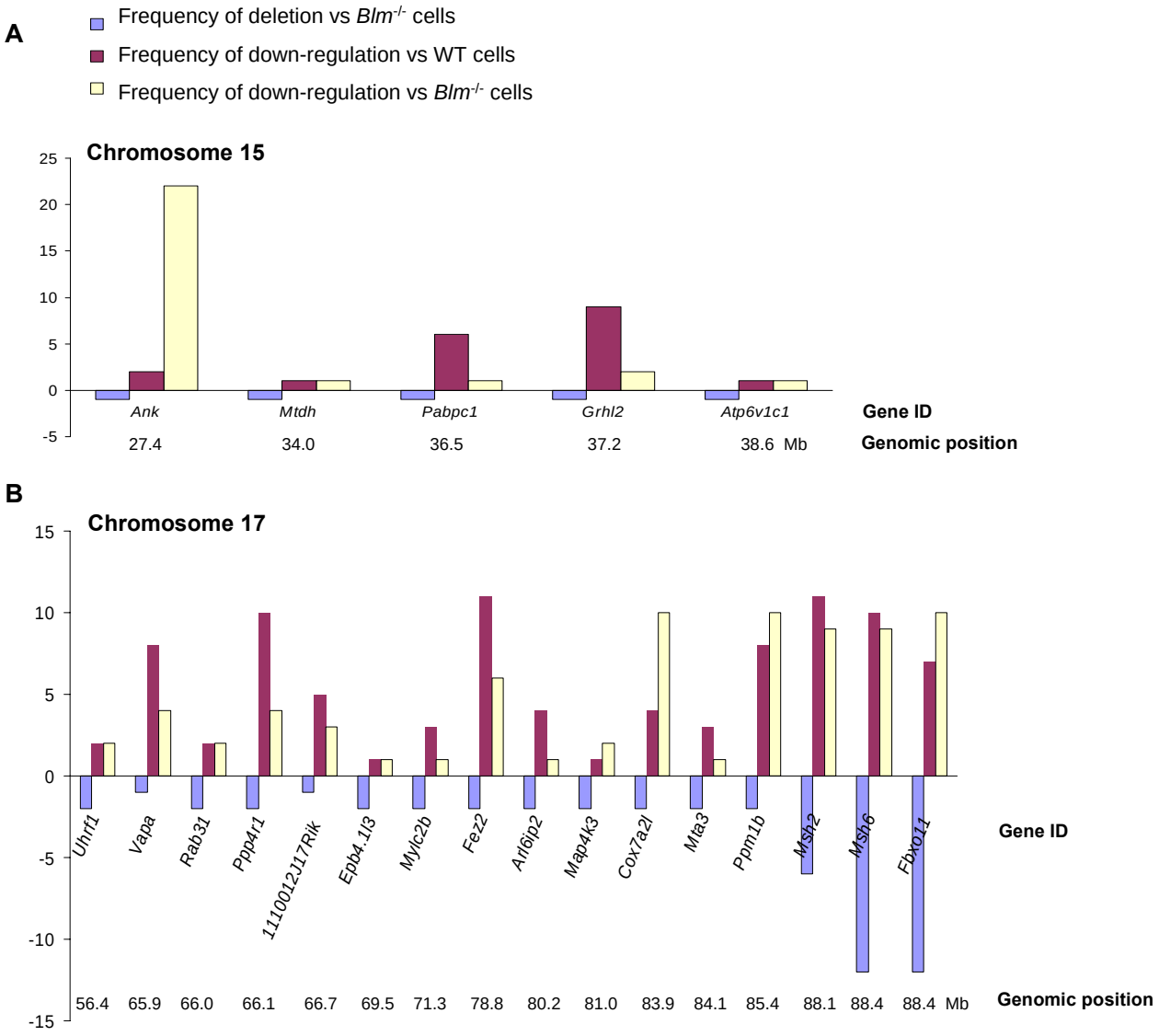


Figure 8-9 Genes on chromosome 15 and 17 which was deleted in one mutant and down-regulated in others

According to Table 8-12, five genes on chromosome 15 and sixteen genes on chromosome 17 were deleted in more than one mutant. And these genes are also down-regulated in more than one mutant compared with the wild type cells (AB1 and AB2.2) and the *Blm*-deficient cells (NGG5-3). These genes are plotted with the frequency of either deletion or down-regulation stated in Table 8-12. These genes are sorted in an ascending order of their genomic position. **A**. Chromosome 15; **B**. Chromosome 17.

8.2.2.3 *Dnmt1* mutant B7

According to array CGH data, gene *Dnmt1* was heterozygously deleted in mutant B7. A 400 kb region (20.4–20.8 Mb) on chromosome 9 is deleted with a Log_2(ratio) of -0.51 (Figure 8-10, a zoomed-in view shown in Figure 8-11). This mutant clone B7 demonstrated strong 6TG resistance, the phenotype of *Dnmt1*-deficiency. It was assumed that there is a small mutation in the undeleted allele of this gene. Therefore, it is worth examining the *Dnmt1* gene at the transcription level to test this assumption. When the expression array data of B7 was compared with the data from the *Blm*-deficient control cells, the transcription signal of the *Dnmt1* probe (scl36157.39.1_18-S) in B7 is 24% of that in the NGG5-3 cell line. This expression decrease can be regarded as a transcript loss. Thus, it can be assumed that in the non-deleted allele there must be some kind of mutation which destroys the *Dnmt1* transcription but which cannot be detected by 200 kb resolution array CGH. The absence of functional *Dnmt1* mRNA will lead to the full resistance to 6TG which was previously observed by Guo (Guo *et al.* 2004).

By comparing the gene expression regulation of the *Dnmt1* mutant B7 with the *Blm*-deficient cells, and comparing that of the *Dnmt1* knockout cell line (ww56+*Hprt*) with its parental wild type cell line AB2.2, six genes (*Fv1*, *Mylpf*, *Rhox2*, *Scml2*, *Sgip1* and *Xlr3a*) were found to have increased expression in the B7 clone and the *Dnmt1* knockout cell line (Table 8-13). Three of them are located on chromosome X and two of them are located on chromosome 4, though their genomic locations are far from each other. While two genes, *Eif2s3y* (eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked) and *Igf2* (insulin-like growth factor 2) were expressed at 10–20% of that in the reference cells. The down regulation of *Igf2* was also seen in the comparison between AB1 and AB2.2 wild type cells (Table 8-1), thus its probe might have high variations. Another explanation is that *Igf2* is expressed in high level in AB1 and it deceased to ~25% in AB2.2 and to ~3% in the two *Dnmt1*-deficient cell lines.

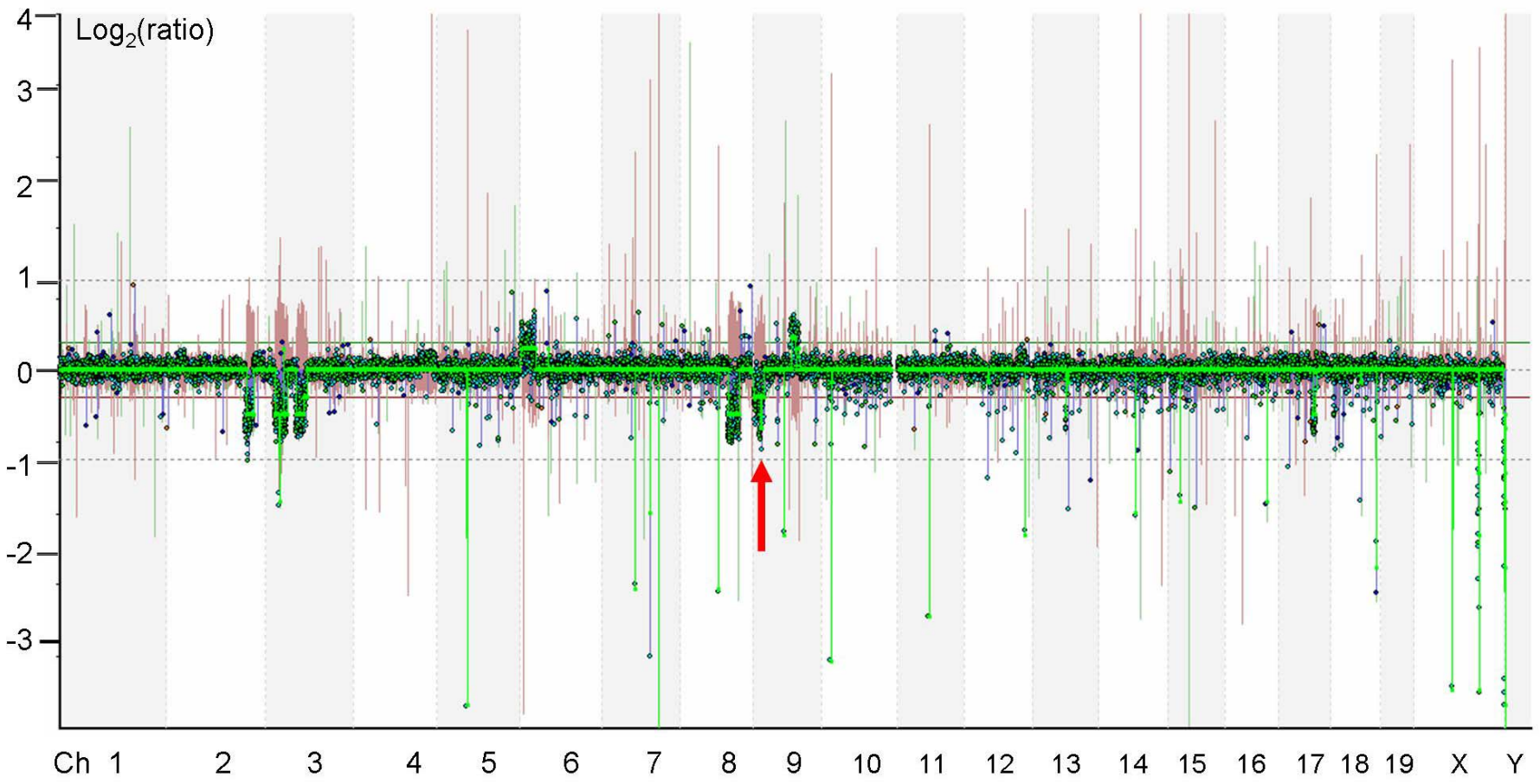
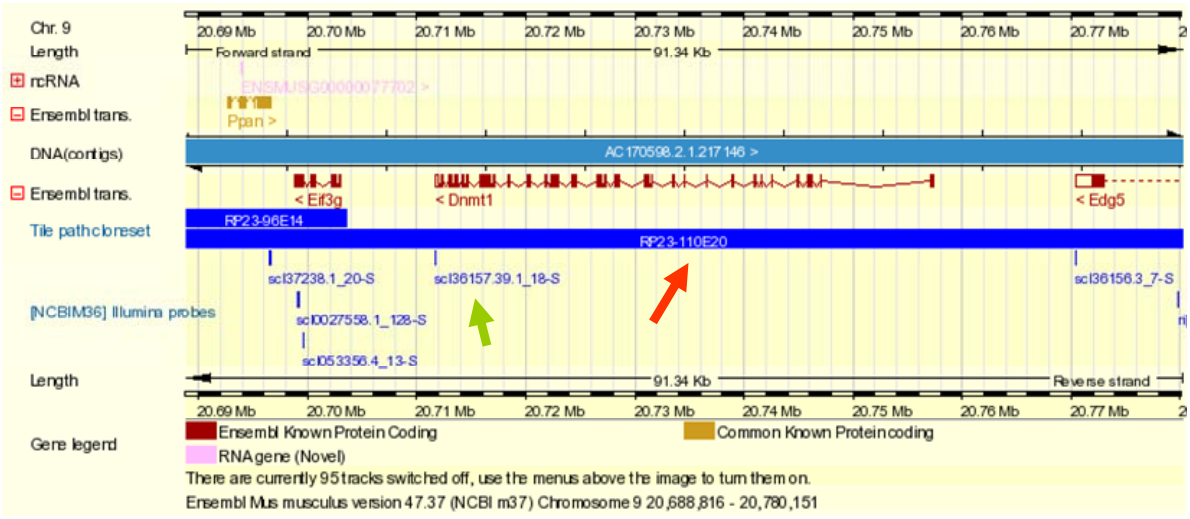


Figure 8-10 **Array CGH profile of mutant B7**

Figure 8-10 Array CGH profile of mutant B7

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 grey dotted lines are for Log_2(ratio) references of either +1 or -1. One deletion (indicated by a red arrow) on chromosome 9 includes the *Dnmt1* locus. The next figure shows a zoomed-in region of the *Dnmt1* locus. There are some single BAC homozygous deletions on chromosomes 3, 5, 7, 8, 9, 11, 12, 13, 17, 18, X and Y. Like this clone, these deletions were found in many other experiments and associated with chromosome Y loss, too. Thus, they were assumed to belong to chromosome Y but incorrectly annotated to other chromosomes. In addition, there are some heterozygous deletions on chromosomes 2, 3 and 8.

A



B

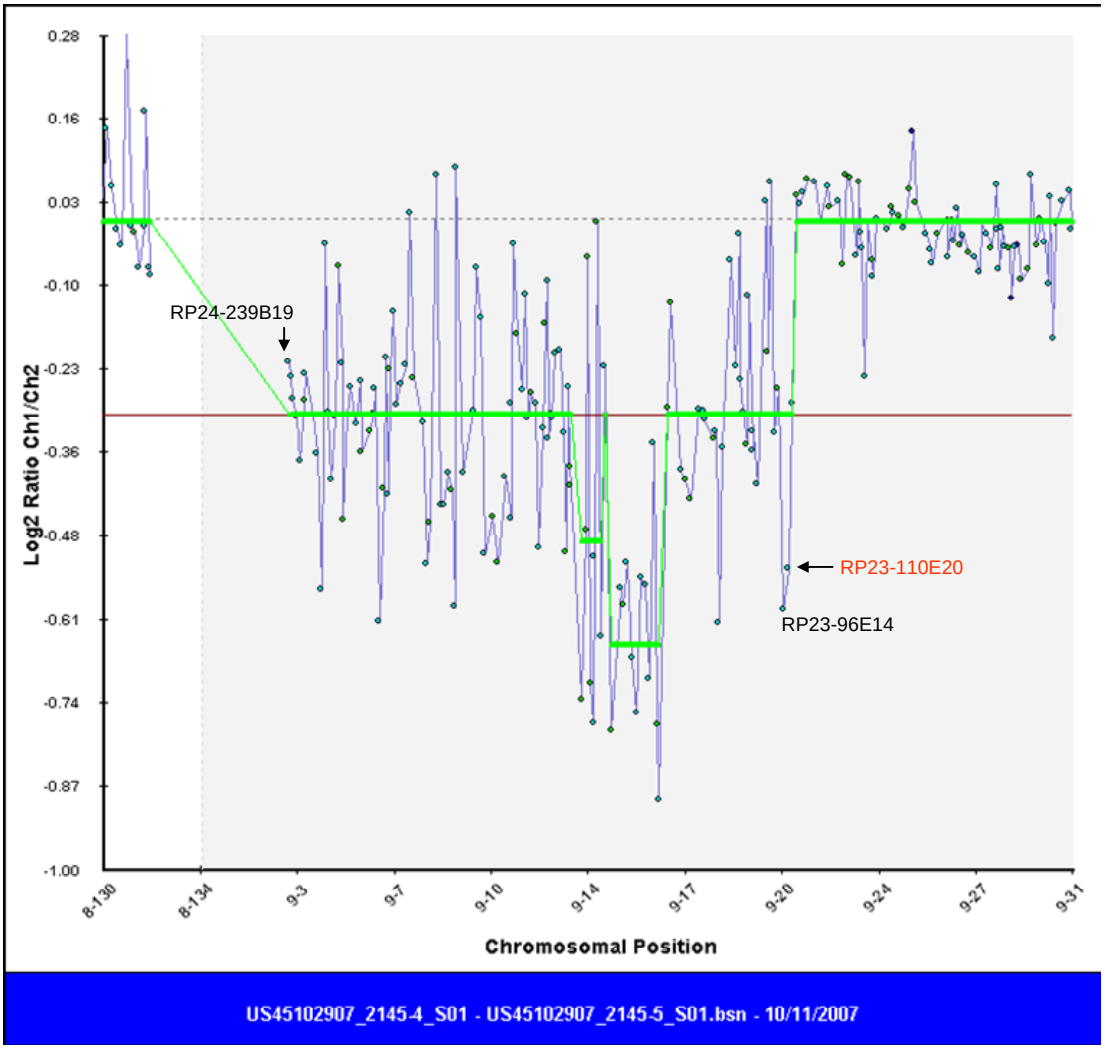


Figure 8-11 **Array CGH analysis of clone B7 at a zoomed-in region of the *Dnmt1* locus**

Figure 8-11 Array CGH analysis of clone B7 at a zoomed-in region of *Dnmt1* locus

A. An Ensembl snapshot of a 91 kb region with the gene *Dnmt1*. Ensembl annotated transcripts are shown in dark red. Tile path BACs for array CGH are shown in blue boxes. BAC RP23-110E20 (indicated by a red arrow) covers the gene *Dnmt1*. Gene-specific probes on the Illumina® array are shown in small blue bars. A probe for the gene *Dnmt1* is indicated by a green arrow. This probe is in exon 39, the last exon of *Dnmt1.* **B**. A zoomed-in array CGH profile of the chromosome 9. BAC RP23-110E20 (red font), covering the gene *Dnmt1,* is within a heterozygous deletion. BAC RP24-239B19 is the starting tile path array BAC on chromosome 9. It is noticeable that the Log_2(Mutant DNA signal/Reference DNA signal) is not the same within a deletion. This can vary depending on the different BAC probes.

Table 8-13 Genes regulated in both mutant clone B7 and *Dnmt1* knockout cells

A

MGI symbol	Description	Expression fold change mutant/reference		Chromosome Name	Gene Start (bp)	Gene End (bp)
		B7	*Dnmt1*$^{-/-}$ knockout			
Fv1	Friend virus susceptibility 1 [MGI:95595]	20.1	3.51	4	147243088	147244467
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle [MGI:97273]	2.89	2.75	7	134355122	134357801
Rhox2	reproductive homeobox 2 [MGI:1922449]	5.43	2.06	X	34784987	34789685
Scml2	sex comb on midleg-like 2 (Drosophila) [MGI:1340042]	4.14	2.01	X	157555125	157696145
Sgip1	SH3-domain GRB2-like (endophilin) interacting protein 1 [MGI:1920344]	6.73	2.07	4	102413902	102644468
Xlr3a	X-linked lymphocyte-regulated 3A [MGI:109506]	2.95	2.17	X	70331634	70342380

B

MGI symbol	Description	% reference signal		Chromosome Name	Gene Start (bp)	Gene End (bp)
		B7	*Dnmt1*$^{-/-}$ knockout			
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked [MGI:1349430]	12%	11%	Y	347055	365035
Igf2	insulin-like growth factor 2 [MGI:96434]	22%	14%	7	149836671	149852721

A. The genes over expressed in the mutant clone B7 and the *Dnmt*$^{-/-}$ knockout cell line (ww56+*Hprt*). **B**. The genes with decreased expression in the mutant clone B7 and the *Dnmt1*-deficient knockout cell line (ww56+*Hprt*). The reference cell of the clone B7 is the *Blm*-deficient cell line (NGG5-3); the reference cell of the *Dnmt1*-deficient knockout cell line is AB2.2 wild type cell line.

8.3 Conclusion and discussion

8.3.1 Confirmation of *Msh2, Msh6* and *Dnmt1* deficiency in mutants

The six mutants, B6, D1, D4, D8, F4, and H14, had homozygous deletions at MMR gene loci, including *Msh2* and *Msh6*. Analysis of these mutant clones with expression arrays confirmed that the transcripts of *Msh2* and *Msh6* genes were significantly reduced. Probe signals were 2.5–25% of levels in the *Blm*-deficient control cells than in the mutants.

Another eight unique mutants (C1, C3, C5, D6, F9, G8, G9 and H13) had lost one allele of the *Msh2* and *Msh6* MMR genes based on array CGH analysis results. Five of these heterozygous mutants were fully resistant to 6TG. Expression array analysis revealed that Four of the five strongly 6TG resistant mutants had lost either *Msh2* (C5 and H13) or *Msh6* (G8 and G9) transcripts (Table 8-8). Their expression array probe signals were 3.8–49 times higher in control cells than in the mutants. This level of expression change was consistent with that observed in the *Msh2* and *Msh6* homozygous deleted mutants and consistent with the strong 6TG resistance phenotype of these four clones. Mutant C3 and D6 were weakly resistant to 6TG and believed to be hypomorphic because they show reduced expression of *Msh2*. In these clones, *Msh2* signals were 40–50% of the level in the controls.

However, the phenotype of heterozygous mutant C1 is not consistent with *Msh2* and *Msh6* gene expression variation. C1 is fully resistant to 6TG, but the transcription levels of both *Msh2* and *Msh6* genes were not significantly reduced. One possible explanation is that a small mutation in the undeleted allele caused amino acid alteration without interfering the normal transcription.

The *Dnmt1* gene was detected as heterozygously deleted in the B7 mutant clone and its transcript was absent, leading to loss-of-function of the Dnmt1 protein. As reported previously (Guo *et al.* 2004), *Dnmt1* deficiency resulted in microsatellite instability and full resistance to 6TG.

By using genomic and expression arrays, a total of eleven mutants were found to have lost-of-function mutations at known MMR genes, including five *Msh2/Msh6* homozygous deletions (B6, D4, D8, F4 and H14), one *Msh6* homozygous deletion (D1), two compound *Msh2* mutations (C5 and H13), two compounds *Msh6* mutations (G8 and G9) and one compound *Dnmt1* mutation (B7). Two out of three mutants (C3, D6 and F9)

with heterozygous *Msh2* and *Msh6* deletions and weak 6TG resistance were examined by expression array analysis. The mutant C3 and D6 showed hypomorphic expression on the *Msh2* gene. One mutant (C1) has a heterozygous deletion at *Msh2* and *Msh6* loci but did not show transcription decrease of these two genes.

8.3.2 Genes regulated by *Msh2* and *Msh6*

The expression array signals of the *Blm*-deficient control cell line NGG5-3 were compared separately with the signals of individual MMR-deficient clones. The expression level of a set of genes (13 genes excluding *Msh6*) were down-regulated to less than 50% in all of the five mutant clones with the *Msh2* and *Msh6* homozygous deletions and in the mutant clone D1 with homozygous deleted *Msh6*. Considering the significance of these data, it is strong evidence that these 13 genes are downstream regulated by MSH6. The normal function of MSH6 is associated with normal transcription of these genes and loss-of-function mutation of MSH6 may play a role in the silencing effect of these genes' expression. Besides, gene *Bst2* was down-regulated in all five *Msh2* and *Msh6* deficient mutants but not in the Msh6 deleted mutant D1. The transcription of this gene is probably downstream of the normal function of MSH2.

In the ten mutants, which lost the transcription of *Msh2* and/or *Msh6* genes, five genes were identified to have 20–50% expression levels of that in the *Blm*-deficient control cells (Table 8-9). These genes (*Col16a1*, *Dcn*, *Lmna*, *Ltbp3* and *Tnfrsf22*) are similar because gene products are extracellular, transmembrane or nuclear membrance associated. The down-regulation of these genes in the MMR mutants may be secondary effects caused by MMR defects.

8.3.3 Potential genes involved in the MMR pathway

By analysing the data from array CGH and expression array, 26 genes (Table 8-12) were found to be homozygously deleted in at least one mutant; at the same time, their transcripts can be observed in the *Blm*-deficient ES cells and the AB1 and AB2.2 wild type cells, but their transcripts were lost in more than two $6TG^R$ mutants. As expected, genes *Msh2* and *Msh6* were among them. Some non-chromosome 17-linked genes such as *Ank, Zfp459* and *Eif5* are in the Table 8-12 with relatively high frequencies of being down-regulated in the $6TG^R$ mutants. This indicates that they are more likely to be involved in the MMR defects. However, the expression decrease of these genes may also be secondary effects of DNA mismatch repair defects.

The selection process in this analysis is stringent as it required genes to be homozygously deleted in one mutant. Considering that compound mutations are common in the previously described *Msh2/Msh6* mutants (Chapter 7), heterozygous deletions may contain MMR genes. Thus, the genes in heterozygously deleted regions also need be investigated in the future.