4 RESULTS

4.1 Gene Dosage Effects in Nine Tissues of a Mouse Model of Down Syndrome

As a first step contributing to understand the molecular basis of DS pathogenesis, we performed a transcriptome analysis of a wide range of tissues in Ts65Dn, an established mouse model of human trisomy 21. Ts65Dn mice have segmental trisomy of mouse chromosome 16 with ca. 128 genes at dosage imbalance (Reeves *et al.*, 1995). The distal boundary of the T(16;17)65Dn chromosome is the most distal gene on mouse Chr16, now known to be *Znf295* (Akeson *et al.*, 2001; Davisson *et al.*, 1990). The proximal breakpoint was previously defined as falling between *Ncam2* and *Gabpa* (Akeson *et al.*, 2001), which are separated by about 3.4 Mb on MMU16 (based on the public mouse genome sequence from Ensembl). To refine this localization, 12 BAC genomic clones that were distributed across the proximal breakpoint region, were isolated from the RPCI-23 library. These were used to narrow the breakpoint region to about 100 kb by FISH analysis of metaphase spreads prepared from Ts65Dn-derived ES cells. The most distal BAC that was present in two copies and absent from the T65Dn chromosome was clone 134F19 (Figure 4-1). The next clone giving a strong positive signal on the T65Dn marker chromosome was 359P19. The breakpoint is thus proximal to *Mrpl39*, as predicted by the elevated expression of three genes proximal to *Gabpa* in trisomic mice (see below).

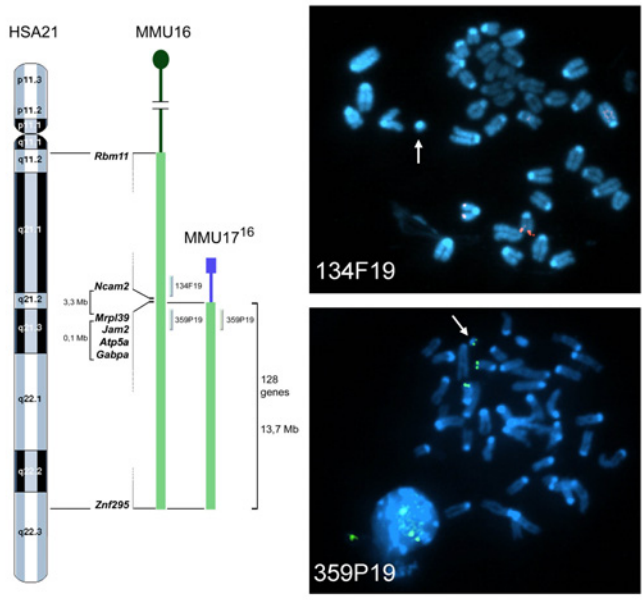


Figure 4-1: Boundaries of the proximal T(16;17)65Dn breakpoint were determined by FISH. Left: Schematic illustrating the syntheny between HSA21 and MMU16, and the translocation chromosome $MMU17^{16}$ with the genes at the boundaries of the triplicated segment. Chromosomal position of the two BACs used for the FISH analysis is indicated. Right: Metaphase spreads from T(16;17)65Dn mouse; BAC 134F19 (top) is in red and BAC 359P19 (bottom) is in green. Arrows point to the $MMU17^{16}$ chromosome.

We designed a mouse cDNA based expression array to interrogate 136 mmu21 genes. Seventy-seven of these are present in three copies in Ts65Dn, from *Mrpl39* to *Znf295* (Figure 4-1). RNA pools from four adult male Ts65Dn mice and four male euploid littermates were prepared from nine tissues (lung, skeletal muscle, midbrain, cerebellum, cortex, liver, testis, heart, and kidney) dissected from three to four month-old mice. The quality of the RNA is essential for the success of the analysis. The integrity of total RNA was verified by agarose-formaldehyde gel electrophoresis (Figure 4-2) for one control sample; for routine verification the RNA quality was evaluated on standard agarose gels, before and after DNase treatment (Figure 4-2). Furthermore the quality of the labeled first strand cDNA was also checked on alkaline gels (Figure 4-3). Probes for which the cDNA bulk smear was lower than 2 kilobase in range were not hybridized.

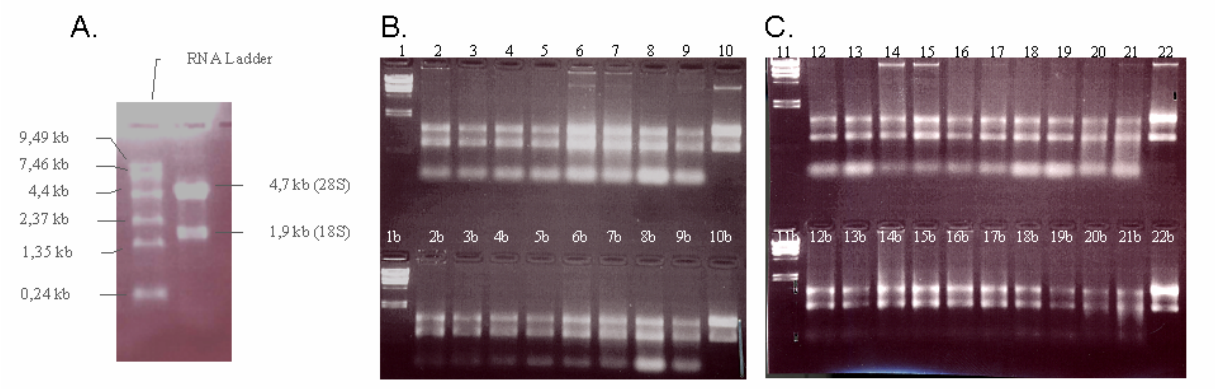


Figure 4-2: RNA quality control. A. Native total RNA on a 1% Formaldehyde gel. *Lane 1:* RNA ladder. *Lane 2:* Total RNA (bands: 28S and 18S ribomal RNA). **B**. Native and DNase treated total RNA on an agarose gel *Lane 1:* λHindIII. *Lane 2-9:* Native total RNA from alternatively Ts65Dn and euploid pools in following order –liver (2,3); testis (4,5); kidney (6,7); skeletal muscle (8-9). *Lane 10*: control RNA (from Hela cells). **C**. *Lane 11:* λHindIII. *Lane 12-22*: Native total RNA from alternatively Ts65Dn and euploid pools in following order cerebellum (12,13); cortex (14,15); midbrain (16,17); heart (18-19); lung (20-21). *Lane* 22: control RNA (from Hela cells). *Lane 1b-22b:* Samples after DNAse treatment, in the same order as in the upper lanes.

Directly labeled first strand cDNA probes from each tissue were hybridized to the arrays in at least two independent hybridizations. Filters were scored semi-automatically and data were analyzed with a dedicated software (Steinfath *et al.*, 2001) (see Material & Methods). A total of 446 non-triplicated (disomic) genes in Ts65Dn mice served as controls. These included 62 mmu21 genes from MMU10, MMU17, and non-triplicated portions of MMU16, plus 384 randomly distributed mouse cDNAs from the Unigene collection (see Material & Methods). The expression profiles of the Chr21 orthologs were compared in euploid and trisomic mice. Data were validated by qPCR for 39 genes.

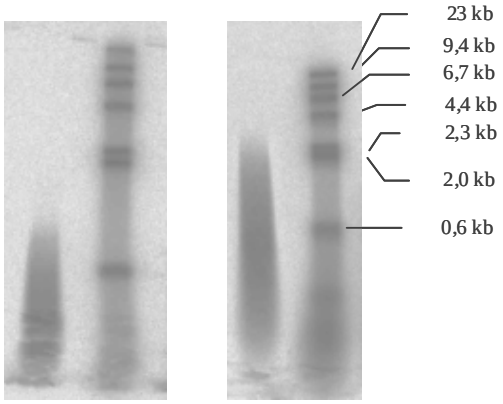


Figure 4-3: Labelled first strand cDNA on an alkaline gel. *Lane 1*: Degraded labelled cDNA, not used for hybridization. *Lane 2*: Marker λ Hind III labelled. *Lane 3*: non-degraded labelled cDNA, used for hybridisation. *Lane 4*: Marker λHind III labeled.

4.1.1 Expression Profiles of Mmu21 Genes in Nine Tissues from Control Mice

Most of the mmu21 genes were found expressed in adult mice. 112 out of 136 mmu21 genes (82%) were detected in at least one of the nine tissues tested. Fifteen out of these 136 genes (11%) were detected in all nine tissues, consistent with a role as housekeeping genes (see full information table at http://chr21.molgen.mpg.de/SI/suppl15231742.html). Skeletal muscle showed the least molecular complexity, where we detected expression for only 15% of the mmu21 genes (Figure 4-4), reflecting its relatively simple cyto-architecture. The most complex transcriptome was seen in brain, which showed expression of 64% of the mmu21 genes (Figure 4-4). The fraction of mmu21 genes expressed in each of the nine tissues is represented in Figure 4-4, showing their functional categories according to Hill *et al.* (Hill *et al.*, 2002). The cerebral cortex showed the highest molecular complexity, indicating that Chr21 genes play an important role in the development and maintenance of this tissue. All functional categories are represented in the cortex and in the midbrain, whereas mmu21 genes coding for ion channels and intracellular signaling molecules are expressed very weakly or not at all in liver and skeletal muscle.

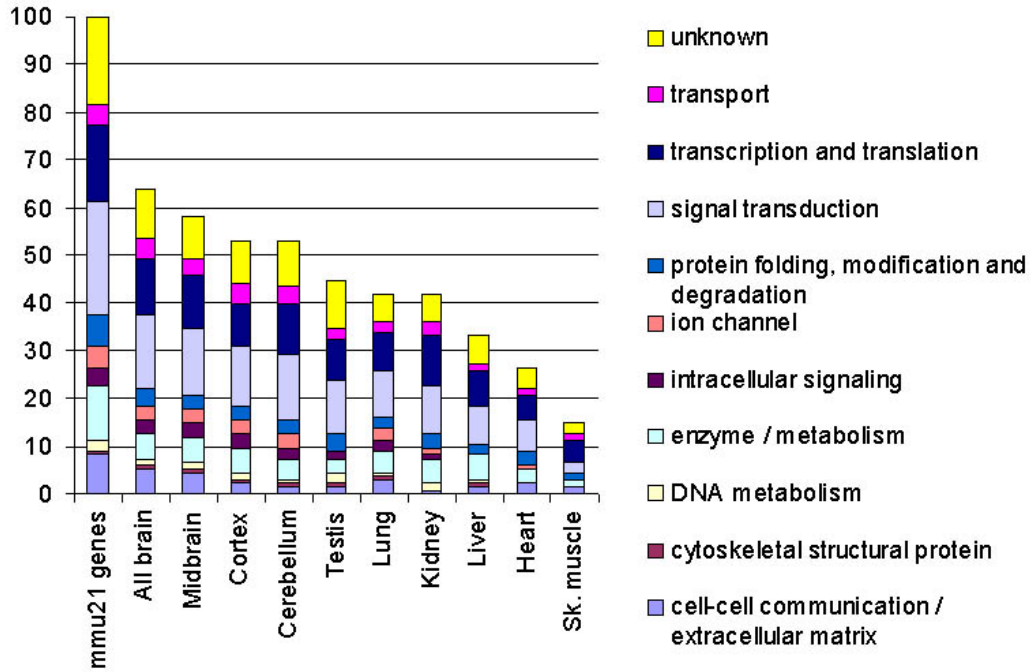


Figure 4-4: Functional distribution of the 136 mmu21 genes in nine tissues of euploid mice. The 136 Chr21 orthologues (mmu21) were distributed into 10 categories of biological processes

(left bar). Bars show the functional distribution of the genes expressed in each tissue, in percent of the total number of mmu21 genes. The right bar (All brain) represents the cumulated data from the three brain regions**.**

Among the 24 genes showing no significant signal on arrays (full information table at http://chr21.molgen.mpg.de/SI/suppl15231742.html), most correspond to relatively rare or spatially restricted transcripts but three, *Znf294*, *Gart* and *Lss*, were shown previously to have widespread expression (Gitton *et al.*, 2002). Eight of these 24 genes have been analyzed by qPCR: *Adamts5*, *Znf294*, *C21orf63*, *Gart*, *Dyrk1a*, *Mx1*, *Tff3*, and *Tmprss3*. qPCR is more sensitive than array hybridization, and all eight genes were detected in at least one tissue (full information table at http://chr21.molgen.mpg.de/SI/suppl15231742.html). However, many of these were only amplified after the 29th cycle, indicating that they are expressed at low levels. Conversely, virtually all genes detected on arrays could be amplified before the 29th cycle (Figure 4-5).

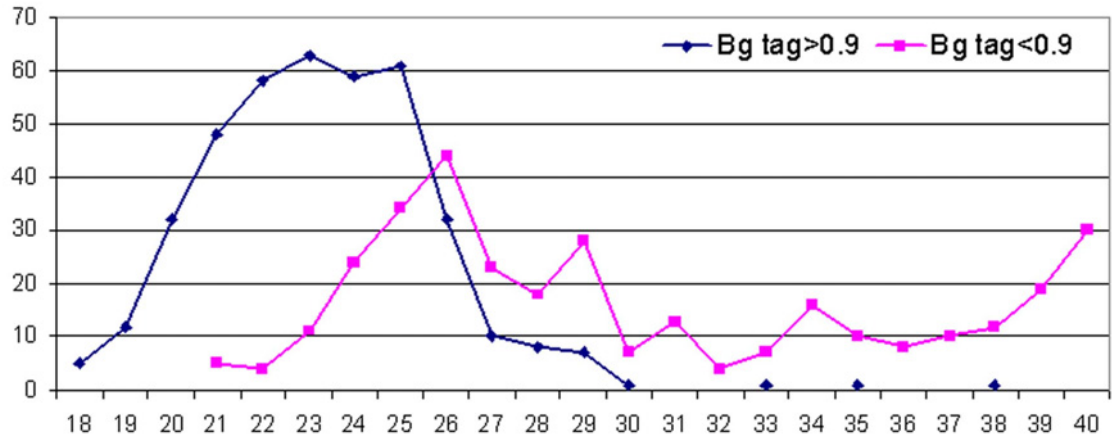


Figure 4-5: Comparison of the expression of genes detected by arrays and Real Time PCR. This graphic combines data corresponding to the genes analyzed by Real Time PCR in all tested tissues. The Y-axis represents the total number of clones detected at a given amplification cycle range (X-axis). Genes that are considered "not expressed" on arrays (background tag < 0.9) are represented by red squares (327 clones); Expressed genes on arrays (background tag > 0.9) are represented by blue diamonds (399 clones).

4.1.2 Molecular Signatures of the Mmu21 Genes in Control and Trisomic Mice Show Tissue Specificity

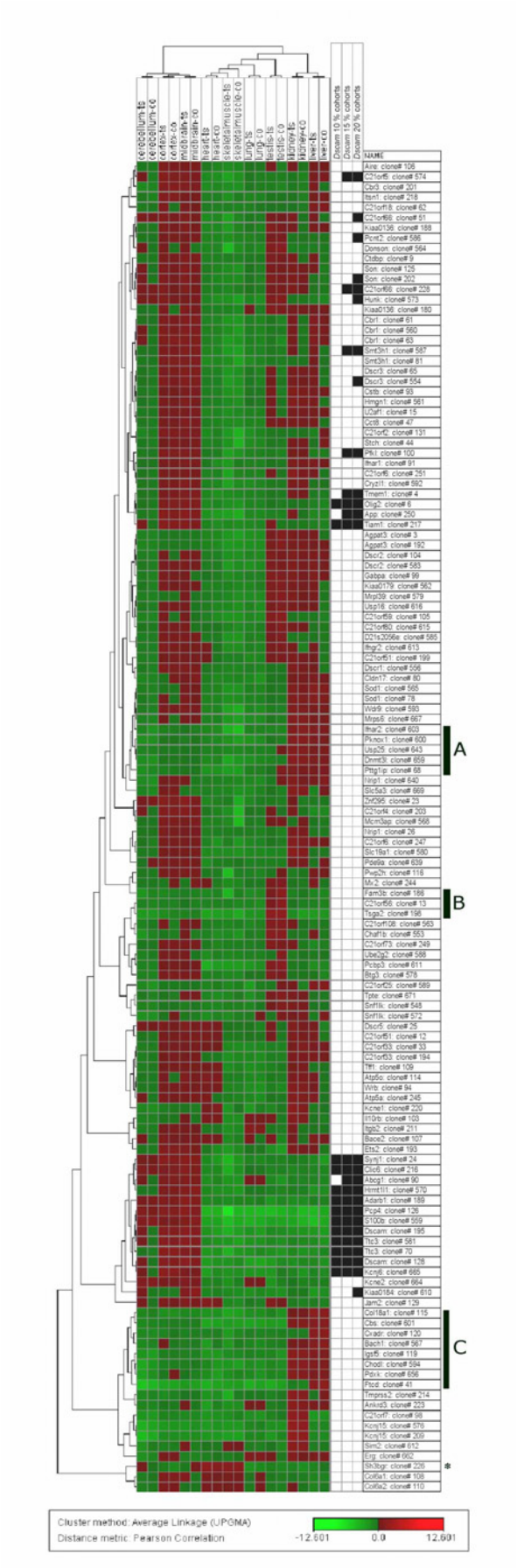
Hierarchical clustering of the array data was used to display the expression patterns of mmu21 genes in nine tissues of the control and trisomic mice. The

relative expression level of a given gene in each tissue was estimated using the average expression value obtained for this gene across all control tissues as a reference (see Material & Methods). The mmu21 genes exhibited characteristic expression profiles in the nine tissues tested here, which could be distinguished and clustered by their molecular signatures. These patterns organized the tissues into several distinct groups (Figure 4-6). For instance, the three brain regions (cortex, cerebellum, and midbrain) clustered together, a feature that was also observed after genome-wide gene profiling (Miki *et al.*, 2001). This might reflect, in part, similarities in their composition (e.g. neurons and astrocytes) and their common embryonic origins. Heart and skeletal muscle also shared significant similarities in patterns of gene expression. A similar order in the clustering of the tissues was obtained by analyzing the expression data of the Unigene control genes (data not shown)

A number of genes were expressed predominantly, though not exclusively, in single tissues. For example, *Sh3bgr* was strongly expressed in heart and skeletal muscle as described previously (Egeo *et al.*, 2000). We observed a group of three genes showing predominant expression in testis, including *Tsga2*, whose gene product is involved in spermatogenesis (Taketo *et al.*, 1997), *C21orf56*, and *Fam3b* that may also play a role in sperm function. Indeed, *C21orf56* and *Tsga2* match ESTs that are derived nearly exclusively from testis (Gitton *et al.*, 2002). Kidney and liver share two significant clusters of highly expressed genes (Figure 4-6), containing several transcripts known to exert a key function in either of these tissues. For instance, *Cbs* catalyses the condensation of homocysteine and serine to cystathionine mainly in liver cells (Kraus *et al.*, 1978), and *Ftcd* is an enzyme linking histidine catabolism to folate metabolism that is mainly expressed in the liver as estimated by EST mining (Gitton *et al.*, 2002). *Col18a1* was reported to be highly expressed in endothelial/epithelial cells of the liver (Lietard *et al.*, 2000), kidney and choroid plexus (Muragaki *et al.*, 1995). *Pttg1ip* is a signal transduction semaphorine-like molecule predominantly expressed in the kidney. Midbrain and cortex displayed very similar profiles of highly expressed genes, many of these being rather weakly expressed in the cerebellum. Among genes predominantly active in the brain, *Dscam* is a cell surface protein acting as an axon guidance receptor (Schmucker *et al.*, 2000). Pearson correlation coefficient identified cohorts

of genes that had expression profiles most similar to *Dscam*. We calculated the 10%, 15% and 20% most similar cohorts (see methods). The ”20%” cohort includes *C21orf5, C21orf66, Pcnt2, Son, Hunk, Smt3h1, Dscr3, Pfkl, Tmem1, Olig2, App, Tiam1, Synj1, Clic6, Abcg1, Hrmt1l1, Adarb1, Pcp4, S100b, Dscam, Ttc3, Kcnj6 and Kiaa0184* (Figure 4-6). As a group, most of these genes are well conserved throughout evolution, are expressed during prenatal development and demonstrate patterned expression in the brain, suggesting their relevance for the development of the central nervous system (Gitton *et al.*, 2002). Several of these (*App, Synj1, Tiam1, Dscam, S100b, Kcnj6, Olig2, Adarb1,* and *Pcp4*) are known to play key roles in brain development and function (Hess, 1996; Leeuwen *et al.*, 1997; Lu *et al.*, 2002; Melcher *et al.*, 1996; Reeves *et al.*, 1994; Saito *et al.*, 2001; Yamakawa *et al.*, 1998). Figure 4-6 furthermore shows that tissues from trisomic mice always clustered with the corresponding euploid tissue, even for the closely matched pairs, demonstrating that transcriptome alterations in Ts65Dn are smaller than the transcriptome differences between tissues. Thus, trisomy appears to predominantly affect levels of expression but not tissue specificity of mmu21 genes.

Figure 4-6 (Next page): Hierarchical clustering showing the expression levels of mmu21 genes across nine tissues of the control mice and Ts65Dn mice. For each clone we calculated the logarithm (base 2) of the ratio between the normalized intensity in the specific tissue and the average of intensities of this clone across the 9 control tissues. Spot intensities below the average of intensities across all tissues give log-ratio values ranging from –12.6 to 0, and are represented by a color gradient spanning from light to dark green. Conversely, spot intensities above the average of intensities across all tissues give log-ratio values ranging from 0 to +12.6, and are represented by a color gradient spanning from dark to light red. 31 clones (24 genes) that did not show significant expression values across all control tissues were excluded from the clustering. 130 clones (112 genes) (rows) and 9 tissues of control and TS65Dn mice resp. (columns) were clustered using the average-linkage hierarchical clustering method with Pearson correlation as similarity measure (J-Express V 2.1, www.molmine.com). Additionally, clones with the most similar expression profiles to *Dscam* (with respect to the Pearson correlation) are displayed: 10%-closest (13 clones, left column), 15%-closest (20 clones, middle column) and 20%-closest (26 clones, right column). Note that in hierarchical clustering procedures, clones with similar expression profiles can be split to different parts of the dendrogram (e.g., *Olig2*), and vice-versa (e.g., *Abcg1*). Genes referenced in the text are highlighted as follows: **Sh3bgr.* Black bars: A: Group of genes highly expressed in the kidney and the liver (from *Ifnar 2* to *Pttg1ip*, p-value = $1.05\ 10^{-4}$) of controls and trisomics. B: Group of genes with predominant expression in the testis of controls and trisomics. C: Cluster of genes highly expressed in the kidney and the liver (from *Col18a1* to *ftcd*, p-value = $5.84\ 10^{-3}$) of controls and trisomics.


cerebellum-ts
cerebellum-co
cortex-ts
cortex-co
midbrain-ts
midbrain-co
heart-ts
heart-co
skeletalmuscle-ts
skeletalmuscle-co
Lung-ts
Lung-co
testis-ts
testis-co
kidney-ts
kidney-co
liver-ts
liver-co
Dscam 10 % cohorts
Dscam 15 % cohorts
Dscam 20 % cohorts
NAME
Aire: clone# 106
C3orf5: clone# 574
Cbr3: clone# 201
Itsn1: clone# 218
C3orf18: clone# 62
C3orf66: clone# 51
Kiaa0136: clone# 188
Pcnt2: clone# 596
Donson: clone# 564
Ctdsp: clone# 9
Son: clone# 125
Son: clone# 202
C3orf66: clone# 228
Hunk: clone# 573
Kiaa0136: clone# 180
Cbr1: clone# 61
Cbr1: clone# 560
Cbr1: clone# 63
Smt3h1: clone# 567
Smt3h1: clone# 81
Dscr3: clone# 65
Dscr3: clone# 554
Cstb: clone# 93
Hmgn1: clone# 561
U2af1: clone# 15
Cct8: clone# 47
C21orf2: clone# 131
Stch: clone# 44
Pfn1: clone# 100
Ifnar1: clone# 91
C3orf6: clone# 251
Cryzl1: clone# 592
Tmem1: clone# 4
Olig2: clone# 6
App: clone# 250
Tiam1: clone# 217
Agpat3: clone# 3
Agpat3: clone# 192
Dscr2: clone# 104
Dscr2: clone# 583
Gabpa: clone# 99
Kiaa0179: clone# 562
Mrp39: clone# 579
Usp16: clone# 616
C21orf59: clone# 105
C21orf80: clone# 615
D21s2056e: clone# 585
Ifngr2: clone# 613
C21orf51: clone# 199
Dscr1: clone# 556
Cldn17: clone# 80
Sod1: clone# 565
Sod1: clone# 78
Wdr9: clone# 593
Mrps6: clone# 667
Ifnar2: clone# 603
Phyox1: clone# 600
Usp25: clone# 643
Dnmt3l: clone# 659
Ptpg1ip: clone# 68
Ninp1: clone# 640
Slc5a3: clone# 669
Zw10: clone# 29
C21orf4: clone# 203
Mcm3ap: clone# 568
Ninp1: clone# 26
C21orf6: clone# 247
Slc19a1: clone# 560
Ptpla: clone# 639
Pwp2h: clone# 116
Mx2: clone# 244
Fam3b: clone# 186
C21orf56: clone# 13
Tsga2: clone# 198
C21orf28: clone# 563
Chaf1b: clone# 553
C21orf73: clone# 249
Ube2g2: clone# 588
Pcbp3: clone# 611
Btg3: clone# 578
C21orf25: clone# 589
Tpte: clone# 671
Snf1lk: clone# 548
Snf1lk: clone# 572
Dscr5: clone# 25
C21orf51: clone# 12
C21orf63: clone# 33
C21orf63: clone# 194
Tff1: clone# 109
Atp5o: clone# 114
Wrb: clone# 94
Atp5a: clone# 245
Kcne1: clone# 220
Ttrb: clone# 103
Itgb2: clone# 211
Bace2: clone# 107
Ets2: clone# 193
Syn1: clone# 24
Clic5: clone# 216
Abcg1: clone# 90
Hmt1l1: clone# 570
Adarb1: clone# 189
Pcp4: clone# 126
S100b: clone# 559
Dscam: clone# 195
Ttc3: clone# 581
Ttc3: clone# 70
Dscam: clone# 128
Kcnj6: clone# 665
Kcne2: clone# 664
Kiaa0184: clone# 610
Jam2: clone# 129
Col18a1: clone# 115
Cbs: clone# 601
Cradr: clone# 120
Bach1: clone# 567
Igsf5: clone# 119
Chst1: clone# 594
Ptrx: clone# 656
Ftcd: clone# 41
Tmprss2: clone# 214
Ankrd3: clone# 223
C21orf7: clone# 98
Kcnj15: clone# 576
Kcnj15: clone# 209
Sim2: clone# 612
Erg: clone# 662
Sh3bgr: clone# 226
Col6a1: clone# 108
Col6a2: clone# 110
A
B
C
*
Cluster method: Average Linkage (UPGMA)
Distance metric: Pearson Correlation
-12.601
0.0
12.601

4.1.3 Global Overexpression of the Mmu21 Genes Located in the Trisomic Region of Ts65Dn

The prevalent hypothesis of gene dosage effect on transcription is that genes located in the chromosomal region present in three copies will be overexpressed by a factor 1.5 fold (Epstein, 1986). We observed a general trend consistent with this idea for most of the triplicated MMU16 genes (Figure 4-7). Assuming that the trend line for diploid genes should be close to one, the level of overexpression was corrected by the value of slope obtained for Unigene controls. The global overexpression was close to 1.5 fold in cortex (x 1.63), heart (x 1.73), testis (x1.51) and liver (x 1.6) whereas somewhat lower values were seen in cerebellum (x 1.37), midbrain (x 1.26), lung (x 1.27), and kidney (x 1.23). Overexpression was collectively less pronounced in skeletal muscle (x 1.16) but was still significantly higher in trisomic than in euploid mice.

To validate the array data by an independent method, we analyzed 39 genes (29 triplicated and 10 duplicated) by qPCR of reversed transcribed RNA from the various tissues. For this we optimized qPCR procedures. Many optimization experiments were done prior to the study (data not shown): optimizing the qPCR conditions, selecting and testing the primers for efficiency and primer dimer formation, setting up routine for qPCR analysis etc.

For the triplicated genes, the general trend of 1.5 fold overexpression was confirmed, apart from a few exceptions discussed below. Duplicated genes generally showed ratios close to 1.0, except for *C21orf56* in testis (x 0.16) or *Tff3* in liver (x 17.8). We observed good concordance with qPCR for 78% of the genes detected on arrays, that is, the expression ratios in trisomic versus controls were found to be within the same range by the two methods (Figure 4-8). Specifically, array ratios with a p-value < 0.05 were generally confirmed by qPCR (see full info table at http://chr21.molgen.mpg.de/SI/suppl15231742.html) Discordant data were most often found for moderately or weakly expressed genes. For instance, most mmu21 genes were weakly expressed in skeletal muscle (Fig. 5), and the difference in expression between trisomic and control mice could not always be established with certainty in this tissue. The analysis was sufficiently precise that elevated expression of *Atp5a*, *Jam2*, and *Mrpl39* in the trisomic mice led to a

refined definition of the translocation breakpoint on the T65Dn chromosome (Figure 4-1 and Figure 4-8).

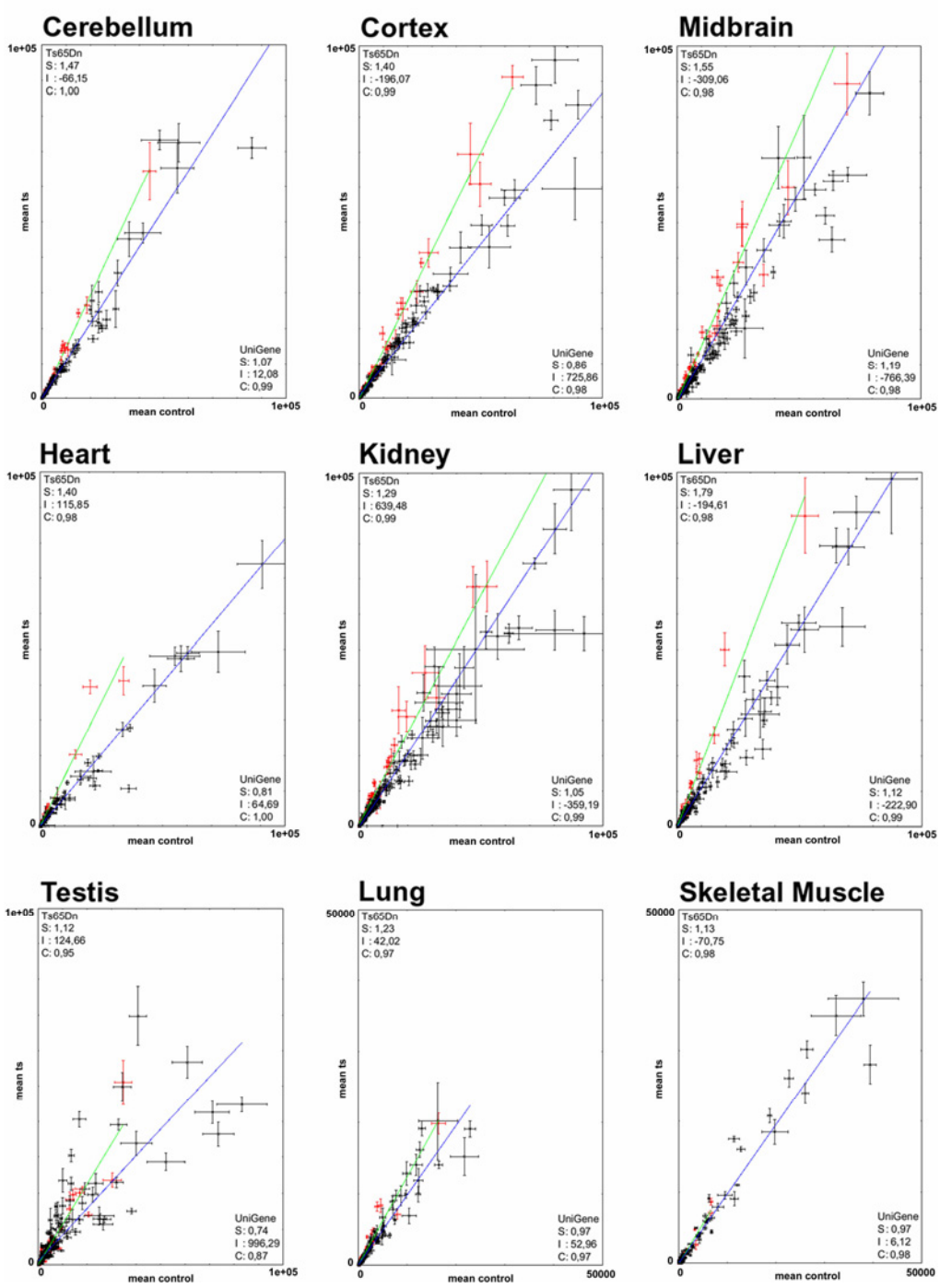


Figure 4-7: Linear regression plots comparing trisomic and controls. For each plot corresponding to a given tissue, the linear regression for the triplicated genes is in green and that of the Unigene sample is in blue. Each clone was plotted using the average of its normalized spot intensities obtained in the replicated hybridizations, with the value in control tissue on the X-axis and with the value in Ts65Dn tissue on the Y-axis. Bars show the interval [ì-ó,ì+ó] where ì and ó are the mean and standard deviation respectively for each data point across the replicated experiments.We excluded from the graphics outlier spots with intensities +/- 4 standard deviations above the mean intensity (e.g. *Wrb*). (S) is the slope for the trendline, (I) the intercept and (C) the correlation. Scales range from 0 to 100.000, except for lung and skeletal muscle, which range from 0 to 50.000.

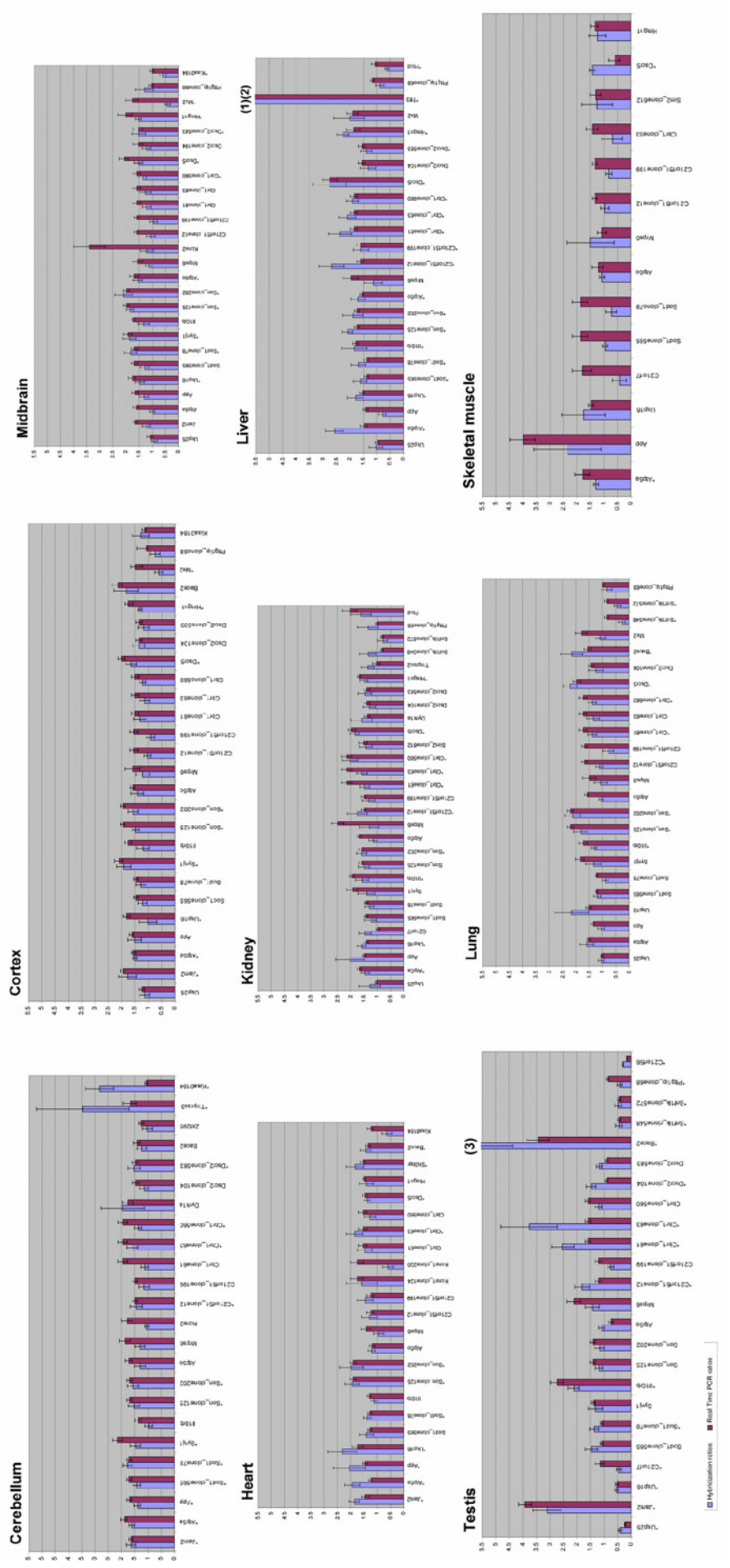

Midbrain
(1)(2)
Skeletal muscle
Cortex
Kidney
Lung
Cerebellum
Heart
(3)
Testis
Hybridization ratios
Real Time PCR ratios

Figure 4-8 (Previous page): Comparison of trisomic/euploid ratios obtained by Real Time PCR and arrays. Histograms showing the correlation between complex hybridization and qPCR results for each tissue. Blue bars represent mean ratios of three independent array hybridizations and red bars represent mean ratios of two independent qPCR experiments. Error bars were calculated by averaging standard deviations of the independent experiments. Genes represented by one to three clones on the arrays were independently compared to the qPCR result. An asterisk (*) beside a gene name indicates an array ratio (blue bar) with significant p-value (<0,05). We found a concordance of 78% between the two methods for the range of expression ratios in Ts65Dn versus controls across all tissues. Truncated bars: (1) ratio = 15.67; (2) ratio = 17.76; (3) ratio= 6.64.

4.1.4 Higher Order Transcript Regulation in Trisomy

Overall, most of the triplicated mmu21 genes were found overexpressed in a range that was proportional to the number of copies of the cognate structural genes. However, a small number of genes escaped this rule. Array analysis identified a few trisomic genes that escaped the “1.5 fold rule” (Table 4-1). These genes were either compensated (transcript levels are not different from euploids), under-expressed, or elevated with a ratio substantially greater than 1.5, in one or more trisomic tissues. These differences were validated by qPCR (see full info table at http://chr21.molgen.mpg.de/SI/suppl15231742.html). We found that nine of the 29 triplicated genes tested by qPCR were either down-regulated or compensated in a single tissue while following the 1.5 fold trend in other tissues (Table 4-1). *Usp16*, a ubiquitin processing protease, and *Atp5o*, a sub-unit of ATP synthase, were down-regulated in testis (x 0.52 and x 0.75). *Dscr5*, a gene of unknown function, was down-regulated in skeletal muscle (x 0.58). *Sh3bgr* was overexpressed in cortex but down-regulated in midbrain (x 0.55). Other genes showed expression ratios close to one and appear to be compensated, suggesting some form of feedback on transcription or other mechanisms controlling transcript levels. *C21orf7*, a putative TGF beta-like activated kinase, and *Tmprss2*, a transmembrane serine protease, were compensated in kidney (x 0.97 and x 1.02). *Dscr2*, *Sod1*, and possibly *C21orf7*, were compensated in testis (x 0.91, x 1.08 and x 1.15). Finally, *Mrps6*, a mitochondrial ribosomal protein, appears to be compensated in skeletal muscle (x 1.07).

Thirteen genes at dosage imbalance displayed a trisomic/euploid ratio > 2 (Table 4-1). Examples are *Bace*2 in testis (x 3.43) an<d in cortex (x 2.15), *Dscr*5 in liver (x 2.75), *App* in skeletal muscle (x 3.99), *Mx2* in heart (x 2.26), *Il10rb* in testis (x 2.73), *Son* in lung (x 2.21), and *Cbr1* in kidney (x 2.13). By qPCR, we also found

enhanced overexpression of *Kcne2* in midbrain (x 3.39), *Mrps6* (x 2.47), *Sh3bgr* (x 2.61) and *C21orf63* (x 3.93) in kidney, *Fam3b* (alias *C21orf11*) in lung (x 2.93), *Mx1* in lung (x 3.7) and in heart (x 2.7). Considering that the triplicated MMU16 segment in Ts65Dn mice contains 128 protein-coding genes, it is possible that as many as 40 genes are indeed compensated or down-regulated in one tissue or another.

Gene	Description	Ratio	Tissue
Usp16	Ubiquitin processing protease	0,52	testis
Atp5o	sub-unit of ATP synthase	0,75	testis
Dscr5	unknown function	0,58	skeletal muscle
		2,75	liver
Sh3bgr	21-Glutamic Acid-Rich Protein	2,61	kidney
		0,55	midbrain
C21orf7	putative TGF beta-like activated kinase	0,97	kidney
		1,15	testis
Tmprss2	Transmenbrane serine protease	1,02	kidney
Dscr2	leucine rich protein C21-LRP	0,91	testis
Sod1	Cu/Zn superoxyde dismutase	1,08	testis
Bace2	aspartyl protease Asp 1	3,43	testis
		2,15	cortex
App	amyloid A4 precursor of Alzheimer's disease	3,99	skeletal muscle
Mx2	human interferon-regulated resistance GTP-binding protein MXB	2,26	heart
Il10rb	human transmembrane receptor protein; cytokine receptor	2,73	testis
Son	SON DNA-binding protein, KIAA1019	2,21	Lung
Cbr1	carbonyl reductase (NADPH) 1	2,13	Kidney
Kcne2	human minK-related peptide 1, potassium channel subunit, MiRP1	3,39	midbrain
Mrps6	mitochondrial ribosomal protein S6	2,47	kidney
		1,08	skeletal muscle
C21orf63	Homo sapiens C21orf63 isoform B protein	3,93	kidney
Fam3b	unknown function	2,93	lung
Mx1	human interferon-regulated resistance GTP-binding protein MXA	3,7	lung
		2,7	heart

Table 4-1: Triplicated genes escaping the 1.5 fold rule: Listed are the gene symbol, a short description of the gene, the Real Time PCR ratio mean [Ts65Dn/Euploid], and the tissue in which the deregulation was observed. Green indicates down-regulation, blues stands for compensation and red for high overexpression.

Although the analysis of genes mapping elsewhere in the genome was not the scope of the present study, several control genes were found with a ratio Ts vs Eu greater or smaller than one, possibly reflecting potential cascade regulations or secondary effects triggered by the overexpression of some of the mmu21 genes. For example, we confirmed by qPCR that two disomic mm21 genes were dysregulated: *c21orf56* in testis (x 0.16) or *Tff3* in liver (x 17.8). Among the 384 Unigene control genes, array data indicated several dysregulated genes, such as mt-Rnr2 (16S ribosomal RNA) up-regulated in testis, Male enhanced antigen-1

down-regulated in testis, Hba-a1 (hemoglobin alpha chain) down regulated in cerebellum and midbrain, and ras-related protein Ral-B, down-regulated in cortex (see Unigene table at http://chr21.molgen.mpg.de/SI/Down_Syndrome/Suppl-Table2.htm).

In order to understand the consequences of trisomy on the genome, we are currently analyzing the whole transcriptome of 8 pooled brain RNA samples of Ts65Dn mice and their euploid littermates using the mouse genome wide expression bead chip from Illumina (http://www.illumina.com/General/pdf/mouse68.pdf). Second, we wanted to investigate further the possible causes of apparent order of gene regulation affecting some of the mmu21 genes. One explanation could be differences in levels of gene expression between individuals, which could not be appreciated in full when using pooled RNA samples.

4.2 Gene Expression Variation in Individual Ts65Dn mice by qPCR

Little is known about the relationship between variation in gene expression levels and phenotypic variation. Identifying those phenotypes that are associated with differences in gene expression is essential to understanding the molecular basis of complex traits and disease susceptibility. Quantitative analyses of gene expression, such as microarrays and qPCR, are frequently performed using pooled RNAs. Pooling schemes mask one important level of information, the natural variation of gene expression (sometimes referred as expression phenotype) among individuals in the population analyzed. Some studies in human or mouse of one or a few trisomic tissues have used RNA from individuals rather than from pools (Mao *et al.*, 2005; Saran *et al.*, 2003), but the issue of inter-individual variation in gene expression has not yet been directly addressed in a comprehensive fashion.

To determine the inter-individual expression differences, we investigated the expression of 50 mouse orthologs of Chr21 in three brain regions in four individual Ts65Dn mice and four control littermates. We performed expression profiling of the mmu21 genes by means of Real Time PCR, allowing the accurate detection of small mRNA differences (estimated here to 1.3 fold) between samples.

Commercially available Taq Man gene expression assays (Applied biosystems) were used.

4.2.1 Expression of Mmu21 Genes in the Brain of individual Mice

RNA was prepared from three brain regions of four trisomic and four euploid young adult male mice. In addition, “biologic-“ or “b-pools” of RNA were prepared by combining equal parts of RNA from the four mice in each genotype class. cDNA from the ten groups (4 Ts65Dn, 4 controls, pool Ts65Dn, pool control) was tested for transcript levels of 50 mmu21 genes by quantitative Real Time PCR in the cerebellum, midbrain and cortex. Of these 50 genes, 33 are triplicated in Ts65Dn whereas 17 genes are located in the segment of mouse chromosome 16 that is disomic in Ts65Dn. Experiments were done in triplicate using two non-mmu21 reference genes previously tested for their stability in each tissue. For each gene, a normalized expression relative to two reference genes (*Hprt* and *Hmbs*) (Appendix Table 7-4, p.153) was calculated. We could reliably detect inter-individual differences as small as 1.3 fold for nearly all of the assays, as the average standard error (SEM) for the relative expression measurements was 0.14 in the 95% confidence intervals.

As mentioned before, a large majority of the mmu21 transcripts were found to be active in the brain. Forty-two genes (31 trisomic, 11 disomic) were expressed in all three of the brain regions tested. Whereas some of those were expressed at high levels in all brain regions, (e.g. *App*, *Son*, *S100b*, *Itsn* and *Dyrk1a*), others were differentially expressed (e.g., higher expression of *Sh3bgr, Col18a1, Tiam1, Pde9a* and *S100b* in the cerebellum, and of *Ncam2* and *Cstb* in the midbrain). Eight genes (*Prss7*, *Cryaa*, *Kcne1*, *Fam3b*, *Tff3*, *Tff2*, *Tmprss3* and *C21orf56*) that had very low or undetectable expression levels (e.g. Ct>34 cycles) were excluded from further analysis.

4.2.2 Gene Dosage Effects in Ts65Dn

We calculated all possible pair-wise combinations of the normalized gene expression values in four Ts65Dn and four euploid mice (Ts/Eu ratios). Thus 16 Ts/Eu ratios per gene were obtained, leading to a total of 2016 ratios for the 42

genes expressed in the three brain regions. Ts/Eu ratios spanned a broad range of values but defined two Gaussian distributions centred around 1.0 and 1.5 for the disomic and trisomic genes, respectively, (Figure 4-9). While the shift between these two distributions clearly demonstrated a global trend of 1.5 fold gene overexpression for the trisomic genes, we observed a substantial overlap in the Ts/Eu ratio distributions. Slight variations around the ideal 1.5 fold value were not unexpected for several reasons, including inherent measurement error and because genetic variation in these mice, which are maintained as an advanced intercross between B6 and C3H, could elicit variable expression phenotypes (see discussion). For this analysis, we considered binned ratios of 0.8-1.2 to be neutral (no change in expression) while binned values from 1.2-2.0 were considered equivalent to 1.5 fold expression change.

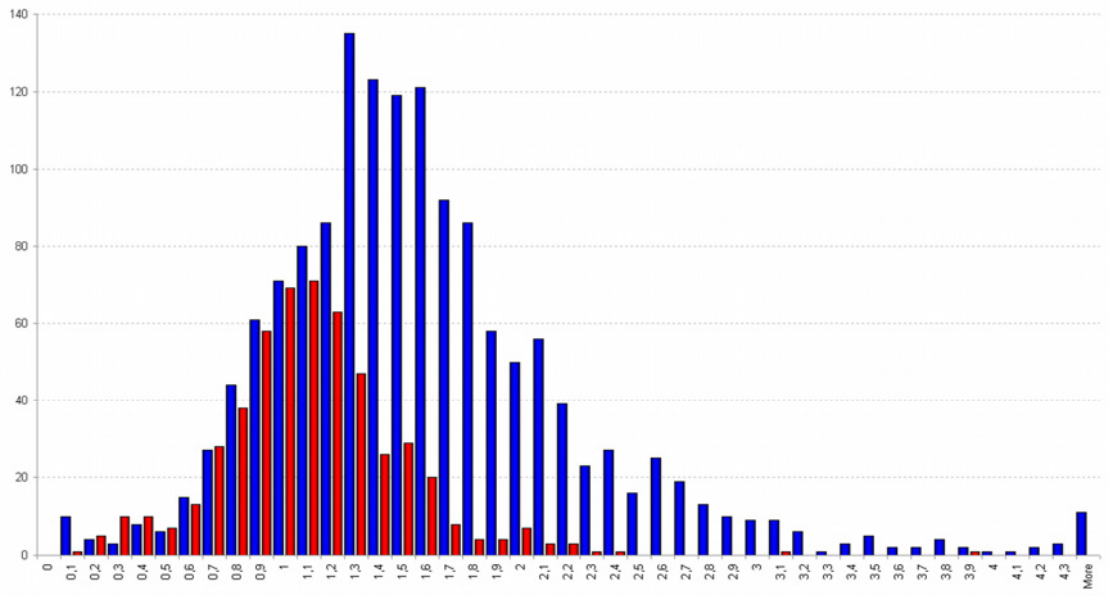


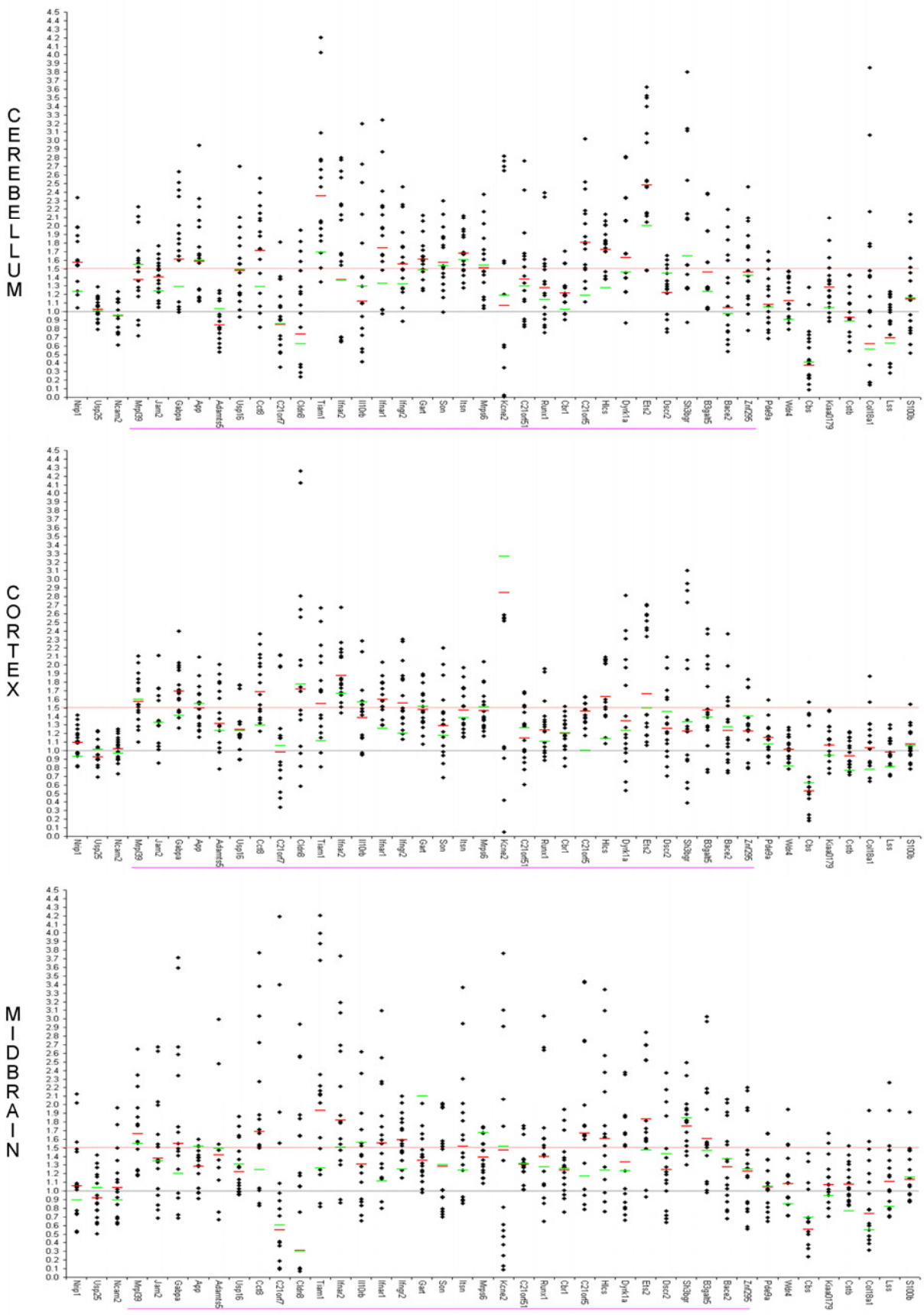
Figure 4-9: Distribution of Ts/Eu relative expression across all tissues.. Histogram of individual pairwise Ts/Eu ratio distribution for all tissues. The red bars represent the distribution of duplicated genes in Ts65Dn, the blue bars the distribution of the triplicated genes. The ratio bins, differing by 0.1, are shown on the X-axis, and the number of values per category is shown on the Y-axis. For each of the brain tissue all pairwise Ts65Dn versus Euploid (Ts/Eu) ratio combinations were calculated corresponding to 16 possible ratios per gene.

Only 53% of all the pair-wise Ts/Eu ratios calculated from individual mice fall within the 1.5 fold-equivalent ranges. In contrast to these individual pair-wise ratio values, which were rather dispersed, 86% of the Ts/Eu ratios obtained for the biological pools were found within the 1.5 fold-equivalent ranges. We plotted all individual Ts/Eu ratios per gene for the cortex, midbrain and cerebellum and

compared these values to the ratios obtained from b-pools of RNA (Figure 4-10). We also calculated the “electronic” pools (e-pools) defined by the ratios of the mean Ts versus mean Eu expression values. We observed a nearly ideal correlation of e-pools with b-pools (Figure 4-10 and Figure 4-11), suggesting that experimental measurement errors are minimal. Pair-wise ratios were somewhat more dispersed in the midbrain (Figure 4-10) than in cortex and cerebellum. For a few genes, we observed drastically dispersed Ts/Eu pair-wise ratios, such as *Kcne2* in all tissues or *Sh3bgr* in the cortex.

Results of mmu21 gene expression in brain regions determined here were in good agreement with those we obtained previously from RNA pools of Ts65Dn and controls using arrays and Real Time PCR. In addition to the three brain regions, we determined here the expression profiles of 24 mmu21 genes (19 trisomic; 5 disomic) in four additional tissues of two Ts65Dn animals and two controls. Averaging the four pair-wise Ts/Eu ratios in e-pools of these four tissues, we confirmed an overall trend of 1.5 fold overexpression of trisomic genes as reported previously for the heart (1.44 $\pm$ 0.05), kidney (1.58 $\pm$ 0.09), liver (1.59 $\pm$ 0.15) and testis (1.36 $\pm$ 0.25) (see Appendix, Table 7-10, p.159). Together with the brain tissues, a total of 132 ratios were compared between the two studies. Of these, we could correlate 74% of the Ts/Eu ratios for pooled RNAs using the binned value ranges described above. Considering that both studies used different mice, that Ts65Dn is not an inbred strain and that the chemistry used for qPCR was different (TaqMan versus SYBR green), the observed correlation is good. A direct comparison could not be made with the study of Ts65Dn gene expression by Lyle and co-workers (Lyle *et al.*, 2004) since they measured expression profiles in the whole brain on RNAs from mice of a different age. However, they also reported an overall overexpression close to 1.5 fold for the trisomic genes in Ts65Dn.

Figure 4-10 (Next page): All Ts/Eu ratios. For each brain tissue and each tested mmu21 genes, we plotted all 16 individual Ts/Eu pairwise ratio (black dots). When values for different individuals of a given poulation were the same, they cannot be distinguish on the graph. The mean Ts/Eu ratios obtained by electronic pooling are represented by the red dashes, whereas the mean Ts/Eu ratio obtained from a biological pool is represented by the green dashes. For *Kcne2*, following pairwise ratios are not visible on the graph: 135.8, 130.1, and 76.8 in cerebellum; 46.3, 45.8, 21.3, 21, 18.7 and 8.6 in cortex; 14.4, 11.2 in midbrain.The X-axis shows each expressed gene in the chromosomal order. Triplicated genes in Ts65Dn mouse are highlighted by the pink line. The 1.0 and 1.5 fold expression values are highlighted by a dotted black and red line respectively.


CEREBELLUM
CORTEX
MIDBRAIN
Ntrp1
Usp25
Ncam2
Mrpl39
Jam2
Gabpa
App
Adamts5
Usp16
Ccl6
C21orf7
Ctsb6
Taam1
Ifnar2
Il10b
Ifnar1
Ifngr2
Gart
Son
Itsn
Mrps6
Kcnn2
C21orf51
Runx1
Cbr1
C21orf5
Hlcs
Dynk1a
Etv2
Dscr2
Slc39a9
B3galt5
Bace2
Znf295
Pde9a
Wbp4
Cbs
Kiaa0179
Cstb
Col18a1
Las
S100b

We previously identified a few trisomic genes that escaped the “1.5 x rule” (Table 4-1) i.e. genes that were compensated (same expression level as euploid) or were highly up- or down-regulated in one or more tissues. Some of these differences were validated here. In the testis, we confirmed that *Usp16* is down regulated (e-pool: x0.44), *Dscr2* is compensated (e-pool: x0.83) and *Bace2* is highly up-regulated (e-pool: x4.23) in trisomic mice. The range of pair-wise individual ratios for *Bace2* fluctuated from 1.2 to 8-fold overexpression in the testis demonstrating substantial individual variation among the tested mice. Three genes that were previously reported not to conform to the 1.5 x rule, *Bace2* in cortex (x2.15), *Kcne2* (x3.39) and *Sh3bgr* (x0.55) in the midbrain, were consistent with the rule in b-pools tested here (Appendix Table 7-11, p.159). However, these genes are among those showing high variability in their expression ratios in at least one tissue (Figure 4-10).

The most extreme case of inter-individual variation in expression was observed for *Kcne2*. Nonetheless, both b-pool and e-pool ratios for *Kcne2* are 1.5 in the midbrain, and do not reflect the individual Ts/Eu ratios which span the largest range (Figure 4-10). Here, another case where the pool ratio is misleading is that of *Bace2* in the cerebellum, which is not different from euploid, suggesting that a compensation mechanism operates on this gene in this tissue. However, analysis of individuals shows a wide range of pair-wise ratio values (0.5 to 2.2), indicating that the pool value misrepresents the actual situation. We observe that genes for which Ts/Eu ratios are skewed in pooled RNAs are most often those with significant expression variation differences between mice.

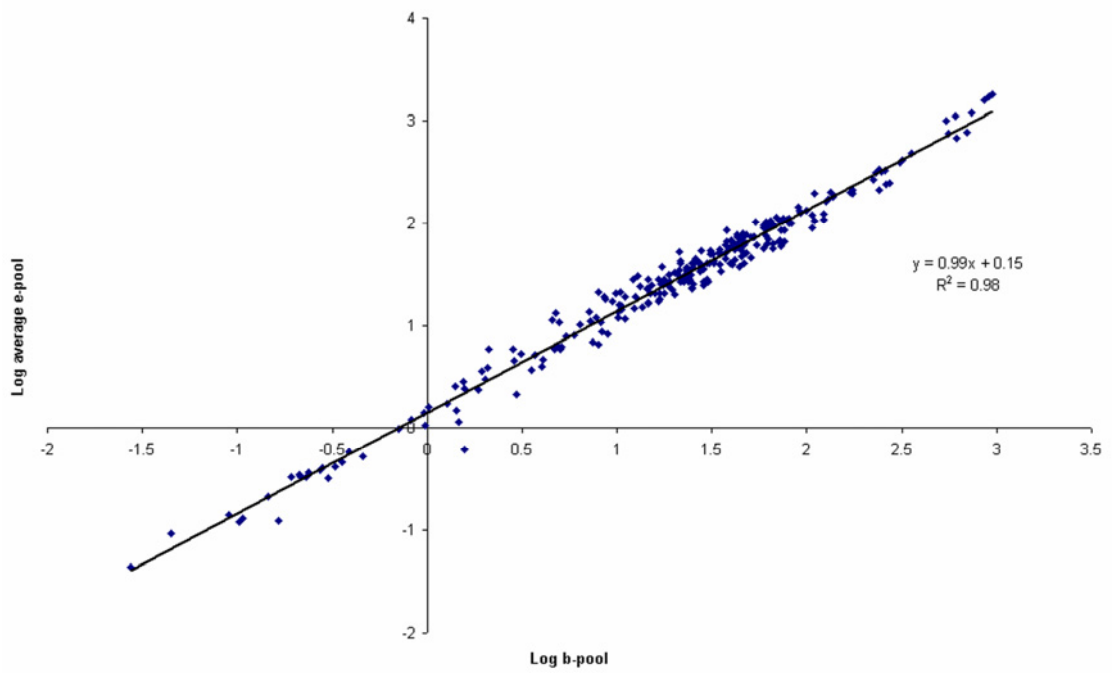


Figure 4-11: Correlation plot of electronic pools versus biological pools. Plot of Log transformed relative expressions of the biological pool (X-axis) against Log of the average relative expressions obtained from electronically pooled data (Y-axis) for each gene (42) in cerebellum, midbrain and cortex. The trendline is shown in black and its equation is displayed on the chart.

Several genes present in two copies in Ts65Dn mice showed altered expression in this analysis. As stated above, *Cbs* RNA levels were reduced in the cerebellum (e-pool: x0.37; b-pool: x0.41), midbrain (e-pool: x0.56; b-pool: x0.69) and cortex (e-pool: x0.53; b-pool: x0.63), and *Col18a1* was reduced in the cerebellum (e-pool: x0.62; b-pool: x0.56). Individual ratios for *Col18A1* were broadly dispersed (0.14 -3.85-fold) in cerebellum, whereas for *Cbs*, there is a trend towards down-regulation (Figure 4-10). In testis, the disomic gene, *C21orf56,* was confirmed to be down regulated (e-pool: x0.18) and in liver, *Tff3* was highly up regulated (e-pool: x2.49) (Appendix Table 7-10, p.159). The individual pairwise ratios for *C21orf56* were consistently down regulated, whereas *Tff3* ratios displayed a broad range of values (0.58-8-fold differences) in liver. In addition, we observed here up-regulation of *Nrip1* in the cerebellum (e-pool: x1.57 and e-pool: 1.23), where pair-wise ratios corroborate pool values. In summary, the data strongly suggest that *C21orf56* is down-regulated and that *Nrip1* is up-regulated in response to trisomy, although we cannot tell if these two disomic genes are directly or indirectly regulated by other mmu21 genes.

4.2.3 Variation of Gene Expression in the Brains of Ts65Dn and Controls

Analysis of individual samples allows the recovery of important information that cannot be determined from pooled samples. We used the coefficient of variation (CV) (see Material & Methods and Appendix Table 7-11, p.159) as an indicator of the variation of gene expression among individual mice and to measure the influence of inter-individual transcript level variation on the Ts/Eu ratios. The CV, in contrast to the variance, is independent of gene expression level and thus enables inter-gene comparison. To estimate the degree to which the technical error influences the biological variation we calculated a technical variance based on the experimental replicates and a biological variance measured between individuals in each group (see Material & Methods and Appendix Table 7-11, p.159). We considered the percentage of technical variance over the total variance (technical + biological variance) as a quality measure. The inter-individual variation in gene expression observed here can be attributed mostly to biological differences since for most genes the biological variance contributes more than 90% of the total variance. Arbitrary cut-offs defined low (<20%), moderate (20-50%) and high (>50%) variation of gene expression according to the coefficient of variation. In control mice, eight genes exhibit a CV<10% (*Jam2*, *Usp25*, and *Itsn* in the cerebellum; *Itsn*, *Mrps6*, *C21orf5*, *Nrip1*, and *S100b* in the cortex; and *App* in the midbrain), suggesting that they are tightly regulated. *App* was reported to show highly variable expression in lymphoblastoid cell lines (Deutsch *et al.*, 2005), but appears to be highly regulated in brain, with a CV of 26% in the cerebellum, 12% in the cortex and 5% in the midbrain. Few genes (3, 5 and 7 genes in the cortex, midbrain and cerebellum respectively) show highly variable expression levels (CV > 50%). Among those, inter-individual *Cbs* expression varies considerably between the three brain regions, corroborating data from lymphoblastoid cell lines (Deutsch *et al.*, 2005). In Ts65Dn mice, 10 triplicated genes had CVs <10 % in at least one brain tissue, e.g., *Cct8*, *Ets2*, *C21orf5*, *Ifnar1*, and *Ifngr2* in the cortex, *Mrps6* and *Usp16* in the midbrain and *Gabpa*, *Hlcs*, and *Ifnar2* in the cerebellum. Conversely, *Kcne2* was found to be highly variable in all three brain tissues, and *C21orf7* in midbrain. In the cortex, only one gene, *C21orf5*, shows a CV <10 % in both Ts65Dn and euploid mice.

For all trisomic genes, we plotted CVs for the Ts65Dn population against the CVs for the euploid population (Figure 4-12). Data showed that 1) for most genes there is no significant difference in the variation of expression between trisomic and euploid mice, since 109 out of 126 data points do not differ by more than 20% in CVs between the two groups; and 2) on average, 90% of the trisomic genes show a low or moderate variation of expression among the eight mice. A few outliers, such as *Kcne2* or *C21orf7*, show dramatic variations of expression among mice.

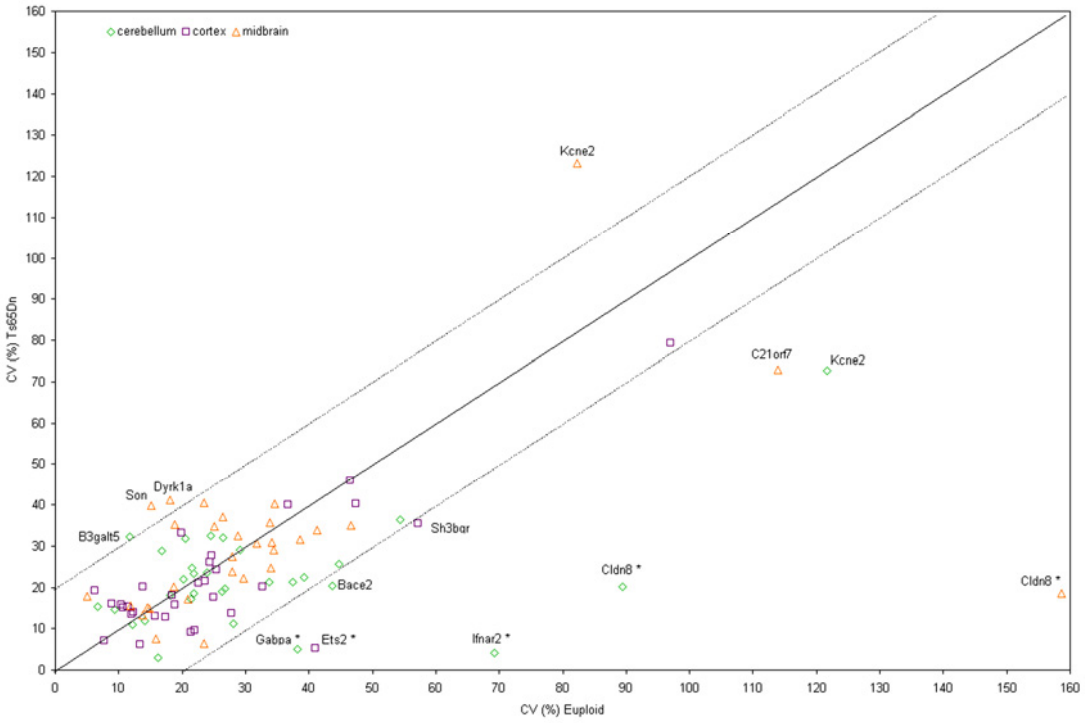


Figure 4-12: Scatter plot of CVs of euploids versus Ts65Dn mice in brain tissues. The coefficient of variation for 31 triplicated mmu21 genes in cerebellum, midbrain and cortex were plotted for Euploid mice (X-axis) against Ts65Dn mice (Y-axis). The dotted lines represent the +/- 20 % CV deviations from the ideal correlation (continuous line). For the data points falling outside this range, the gene name is added.

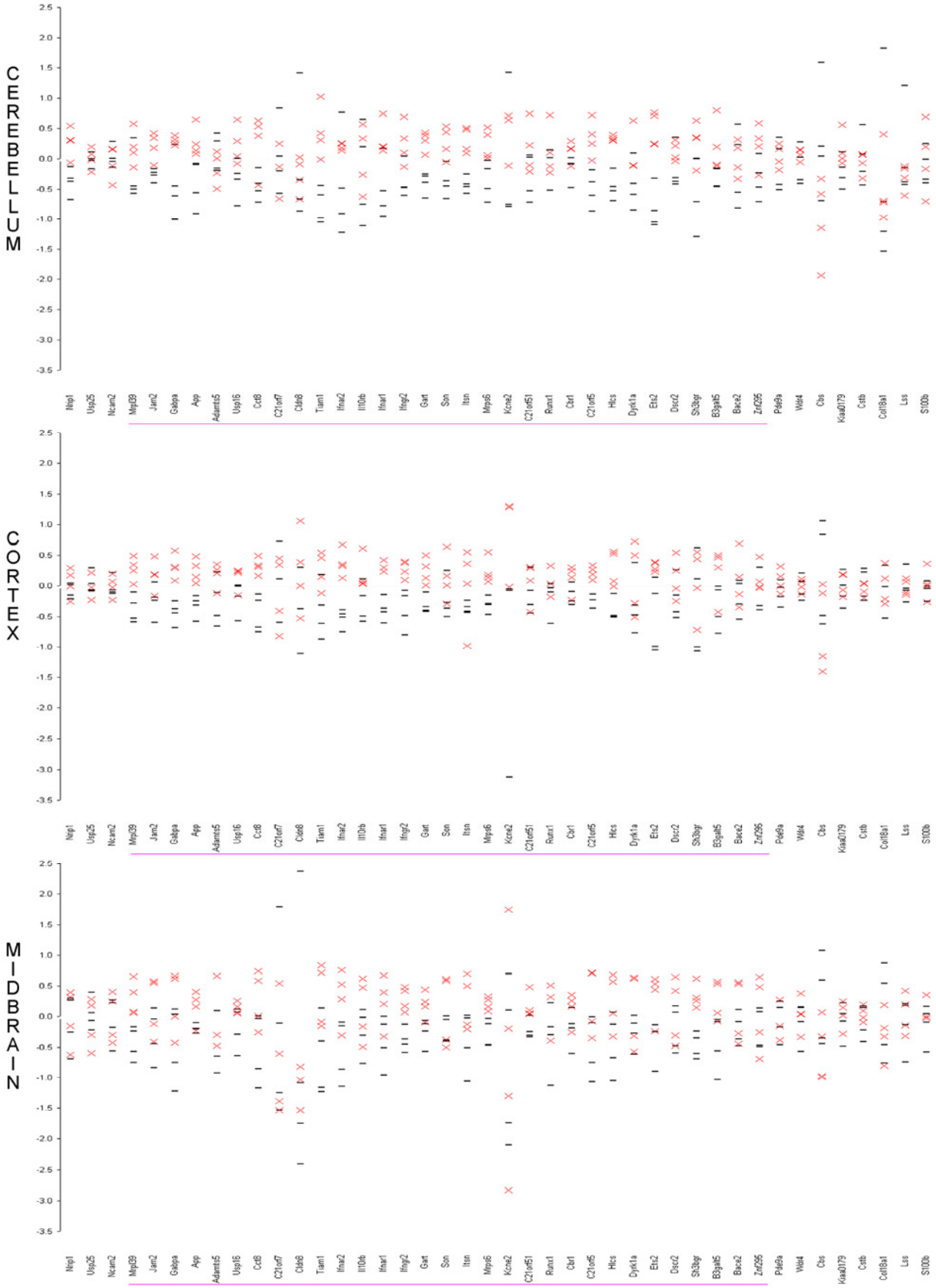
To determine statistically whether the presence of an extra copy of a gene influences its variation in expression level, we used the F-Test to assess whether the two populations differ significantly by their variance (Appendix Table 7-11, p.159). It should be noted that the variance of the expression levels increases as the gene expression increases. This obscures the fact that in proportion the level of variation is not greater. We used the coefficient of variation calculated with the F-test to circumvent this artifact.

Overall, the amplitude of variation of gene expression did not differ significantly between euploid and Ts65Dn mice. The basal level of expression of a given gene follows a Gaussian distribution of similar width in trisomic and in euploid mice. This result confirms earlier findings in fetal cortex of human trisomy 21 (Mao *et al.*, 2003), at least as far as mmu21 genes are concerned. Contrary to expectations, however, the coefficient of variation was significantly smaller in trisomic mice for four genes in cerebellum (*Gabpa,* F-test, p=.0.007; *Cldn8*, F-test, p=0.03; *Ifnar2*, F-test, p=0.001; and *Hlcs*, F-test, p=0.02), for *Ets2* in cortex, (F-test, p=0.01) and for Cldn8 in midbrain (F-test, p=0.005).

4.2.4 Prioritization of Candidate Genes for the Trisomy Phenotypes

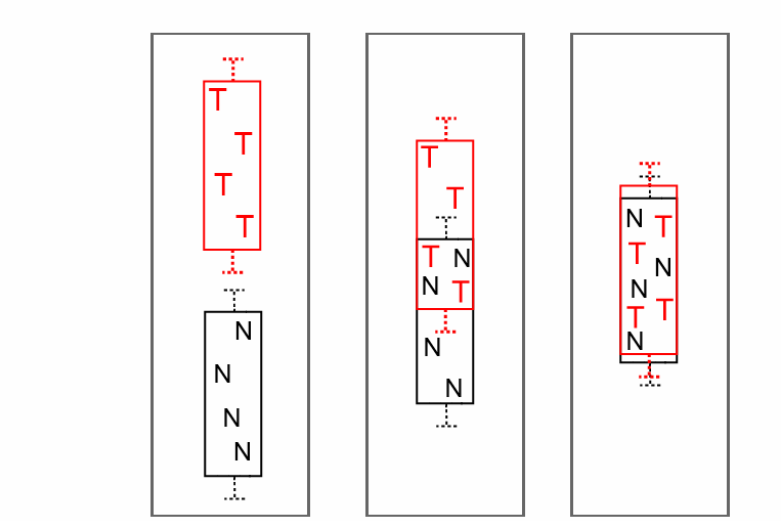
We stratified here trisomic genes based on their individual expression profiles in the brain. To represent the expression levels of mmu21 genes (disomic and trisomic) among individual mice, we plotted the log ratios of the individual normalized gene expression values over the mean expression across all the eight mice (Figure 4-13). Three categories of genes can be distinguished from the normalized expression levels: 1) genes whose expression levels in the four Ts65Dn mice are clearly above that of their euploid littermates (e.g. *App* in the cortex); 2) genes whose expression levels are partially overlapping between the two populations of mice (e.g. *Jam2* in the cortex); and 3) genes whose expression levels overlap substantially between the two populations and thus cannot be used to distinguish Ts65Dn from euploid (e.g. *Ncam2* in the cortex). As expected, all of the 11 disomic genes tested here were found in the latter category, except *Nrip1*. The nuclear receptor co-repressor *Nrip1* was found to be significantly up regulated by an average of 1.57 fold (p=0.01) in the cerebellum of Ts65Dn mice, suggesting that *Nrip1* might be regulated by a trisomic mmu21 gene.

Figure 4-13: Relative expressions plot (Next page). For each four Ts65Dn mice (*red crosses*) and four Euploid mice (*black dashes*) in cerebellum, cortex and midbrain the Log2 ratio of the individual normalized expression over the mean expression across all individuals was plotted. When values for different individuals of a given poulation were the same, they cannot be distinguish on the graph. On the X-axis each gene with significant expression in all sample is represented in the chromosomal order. Triplicated genes in Ts65Dn mouse are highlighted by the pink line below the gene's name. Following values are not visible for *Kcne2* on the cerebellum plot [-6.38 (euploid); -6.12 (Ts65Dn)] and on the cortex plot [-4.24 (euploid); -4.36 (Ts65Dn)].


CEREBELLUM
CORTEX
MIDBRAIN
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
-3.5
Ntrp1
Usp25
Ncam2
Mpd39
Jam2
Galpa
App
Adamts5
Usp16
Ccd8
C21orf7
Cldn8
Tiam1
Ifnar2
Il10rb
Ifnar1
Ifngr2
Gart
Son
Itsn
Mrps6
Kcne2
C21orf51
Runx1
Cbr1
C21orf5
Hlcs
Dyrk1a
Ets2
Dscr2
Sh3bgr
B3galt5
Bace2
Znf295
Pde9a
Wdr4
Cbs
Kiaa0179
Cstb
Col18a1
Lss
S100b

Based on the distribution of expression levels in the brain, the three categories of mmu21 genes at dosage imbalance are presented in Figure 4-14. The first category contains genes whose expression in an individual trisomic mouse was always significantly higher than in any euploid animal tested ($p<0.05$). It should be noted that in this first category, we observed the case of a few genes for which one individual out of the trisomic or control group show slight overlap of its transcript level with that of the second group, still associated with a consistent and significantly elevated expression levels in Ts65Dn compared to euploid mice ($p<0.05$) (e.g *Gabpa* in the cerebellum, Figure 4-13). Across the three brain tissues, nineteen genes show expression levels significantly higher in trisomic than in euploid mice.

We speculate that genes in this first category may have a greater penetrance in the cerebellar phenotypes observed in mouse models of Down syndrome (O'Doherty *et al.*, 2005; Olson *et al.*, 2004; Roper *et al.*, 2006) and may also be important candidates in structural and functional deficits in the DS brain. Notably, *App*, *Ets2*, *Gart*, *Ifngr2* and *Mrps6* belong to this highly differentiated category for the three brain regions. Further, many genes of this category are conserved at least in *C. elegans* and *Drosophila* and are tightly regulated with a CV < 10% in trisomic, euploid or both, pointing out to genes that are important for the organism development (*Jam2*, *App*, *Cct8*, *Itsn*, *Mrps6*, *C21orf5*, *Ets2*), with the exception of the interferon receptors *Ifnar*1 and *Ifnar*2 which show a low CV but are not conserved. *Ets*2 is a transcription factor involved in a number of processes and is a key regulator of immunity (Gallant and Gilkeson, 2006). Two genes that are good candidates for neurodegenerative pathologies associated with DS are *App* mutated in some forms of Alzheimer disease (Selkoe, 1997), and Intersectin or *Itsn* involved in clathrin-mediated endocytosis (Keating *et al.*, 2006). *Jam2* is associated with cell adhesion processes and *Cct8* with protein folding and degradation. *C21orf5* is a new member of the Dopey family and exhibit restricted regional brain expression (Gitton *et al.*, 2002).



Cerebellum	*Jam2* *Gabpa* *App* *Cct8* *Tiam1* *Ifnar1* *Ifngr2* *Gart* *Son* *Itsn* *Mrps6* *C21orf5* *Hlcs* *Ets2*	*Mrpl39* *Adamts5* *Usp16* *Ifnar2* *C21orf51* *Runx1* *Cbr1* *Dyrk1a* *Dscr2* *Sh3bgr* *B3galt5* *Znf295*	*C21orf7* *Cldn8* *Il10rb* *Kcne2* *Bace2*
Cortex	*Mrpl39* *Gabpa* *App* *Cct8* *Ifnar2* *Ifnar1* *Ifngr2* *Gart* *Itsn* *Mrps6* *C21orf5* *Hlcs* *Ets2*	*Jam2* *Adamts5* *Usp16* *Cldn8* *Tiam1* *Il10rb* *Son* *Kcne2* *Runx1* *Crb1* *Dyrk1a* *Dscr2* *B3galt5* *Znf295*	*C21orf7* *C21orf51* *Sh3bgr* *Bace2*
Midbrain	*Mrpl39* *App* *Ifnar2* *Ifngr2* *Gart* *Mrps6* *C21orf51* *Ets2* *Sh3bgr* *B3galt5*	*Jam2* *Gabpa* *Adamts5* *Usp16* *Cct8* *Tiam1* *Il10rb* *Ifnar1* *Son* *Itsn* *Runx1* *Crb1* *C21orf5* *Hlcs* *Dyrk1a* *Bace2*	*C21orf7* *Cldn8* *Kcne2* *Dscr2* *Znf295*

Figure 4-14: Gene categorization to phenotype penetrance. Genes are gouped in 3 categories according to their p-value in the t-test and to the tissues were they were tested. The first category (left) shows genes with $p<0.05$, meaning that the expression levels in Ts65Dn individual mice are consistently different from euploids. The second category shows genes with $0.05<p<0.34$, for which the expression levels of Ts65Dn samples partially overlap with euploids. The last category ($p>0.34$) groups the genes for which the expression levels between Ts65Dn and euploid mice cannot be distinguished

The second class represents the genes for which the expression overlaps partially between trisomic and euploid mice (p-values between 0.05 and 0.33). In

general, genes in this category vary moderately and ¾ of the CVs of Ts65Dn and euploids are above 20%. We hypothesize that in general, genes whose expression is more variable are more likely to contribute to those pathogenic outcomes of DS that show variable expressivity.

The third category is populated by a few trisomic genes for which expression levels are intermingled between the two populations of mice (p-values> 0.33). They all have moderate to high expression variation based on CV. Thus, they may be less likely to contribute to trisomy-related phenotypes. Genes whose expression levels vary widely among individuals are more likely to reach a critical threshold of overexpression in a subset of individuals, and thus might be more likely to be involved in phenotypes that occur in some but not all individuals with trisomy 21.

We have now extended the analysis to eight trisomic and eight control mice. We essentially confirm the observations made here.

4.3 Functional Analysis of Chr21 Orthologs in *Caenorhabditis elegans*

Although the systematic analysis of mmu21 gene expression profiles contributes to an understanding of how cells and organisms respond to structural gene dosage imbalance, it does not give direct information on gene function. Thus, nine genes out of the 89 clear orthologs of HSA21 in the model organism *Caenorhabditis elegans* have been selected for their potential relevance in DS phenotypes and their orthologs have been functionally investigated in the worm. To do so, we monitored the *in vivo* expression of the selected genes under the control of their endogenous promoter using GFP fusions. To further track the processes where these candidates could be implicated in, we knocked-down individually their expression in three different strains of *C. elegans* (*N2, VH525, rrf-3*) by RNAi. An overview of the results is provided in Table 4-2.

HSA21 gene name	Gene Aliases	Description	Gene Ontology	Ensembl GeneID	Gene ID (NCBI)	Longest protein length (Amino Acids)	expression profile results in Ts65Dn	Ortholog transcript ID (WB)	Chr. location	Protein ID(WB)	Longest protein length (AA)	E-value	% length	Expression GFP fusion	RNAi Screen
CBS	HIP4	human cystathionine-beta-synthase	enzyme / metabolism essential enzyme for the removal of homocysteine	ENSG00000160200	875	565	ratio ~ 1 in cerebellum, cortex, kidney, liver and midbrain (it a duplicated gene in Ts65Dn). Not detected in the other tissues	ZC373.1	X	WP:CE31670	704	5.5E-93	46.30%	Cytoplasmic, intestine, head and bodywall muscles	Wild type
DIP2A	C21orf106, Dip2, KIAA0184	human mRNA for KIAA0184 protein	unknown	ENSG00000160305	23181	1573	ratio ~1 in all tissues except in testis (1.4). Not detected in skeletal muscle.	F28B3.1	I	WP:CE40204	1539	4.2E-280	74.70%	Not available	Wild type
HEMK2	C21orf127, MGC19995, MTQ2, N6AMT1, PRED28	HemK methyltransferase family menber 2	nucleic acid binding / DNA replication and modification	ENSG00000156239	29104	214	not reported	C33C12.9	II	WP:CE27815	234	5.8E-39	80.30%	Neuronal (ventral nerve chord, head neurons, commissures)	Wild type
TMEM50B	C21orf4, DKFZp686C2482, HCVP7TP3	transmenbrane protein 50B	unknown	ENSG00000142188	757	158	~1.5 overexpression in all tissues, but not detected in heart, liver, skeletal muscle.	Y74C10AL.2	I	WP:CE21753	157	6.7E-38	98.70%	cytoplasmic, seam cells, hypodermis, few neurons	Wild type
ABCG1	ABC8, MGC34313, WHITE1	white protein homolog (ATP-binding cassette transporter 8)	transporter / transport	ENSG00000160179	9619	825	ratio ~ 1 in cerebellum, cortex, lung and midbrain (it a duplicated gene in Ts65Dn). Not detected in the other tissues	C05D10.3	III	WP:CE29170	598	1.80E-76	98.20%	No expression detected	Not available
ITSN	ITSN, MGC134948, MGC134949, SH3D1A, SH3P17	human intersectin-SH3 Domains-containing protein SH3P17	signal transduction / multifunctional proteins	ENSG00000205726	6453	1721	~1.5overexpression in cerebellum, cortex and midbrain. Not detected in othr tissues	Y116A8C.36	IV	WP:CE23342	1097	3.6E-106	94.70%	Not available	Wild type
SH3BGR	SH3 domain binding glutamic acid-rich protein	21-Glutamic Acid-Rich Protein (21-GARP)	signal transduction / intracellular signaling	ENSG00000185437	6450	239	~1.5 overexpression in cerebellum, cortex, heart and lung. In kidney (2.61) overexpression, In midbrain (0.55) downregulation, compensated in skeletal muscle (0.99), not detected in testis.	Y105E8A.1	I	WP:CE36221	132	7.30E-13	72.70%	cytoplasmic, bodywall muscle, pharinx, head muscles, sex muscle, anal sphincter and depressor muscle, few head neurons.	Not available
DONSON	B17, C21orf60, DKFZP434M035	downstream neighbor of SON	unknown	ENSG00000159147	29980	566	~1.5 overexpression in cerebellum and cortex; seems downregulated in testis (0.65), and is not detected in the other tissues	C24H12.5	II	WP:CE08359/ WP:CE27729	612	4.30E-08	48.90%	Not available	Wild type
KIAA0179		human mRNA for KIAA0179 protein	unknown	ENSG00000160208	23076	758	cerebellum(1.74) and cortex(1.47), and testis (0.86). Not detected in the other tissues.	C47E12.7	IV	WP:CE20571	397	2.4E-29	96.50%	Not available	Mig/Gro/Lva

Table 4-2 (Previous page): HSA21 candidates and their orthologs in C.elegans. *Column 1*: Human chromosome 21 (HSA21) gene name. Genes duplicated in the Ts65Dn mouse (green), genes triplicated (red) in the Ts65Dn mouse. *Column 2*: alternative gene names. *Column 3:* gene description. *Column 4*: gene ontology. *Column 5*: Ensembl gene accession number. *Column 6:* Entrez gene accession number. *Column 7:* amino acid length of the longest protein encoded by the human gene. *Column 8*: summary of the expression profile observed in the Ts65Dn mice study. *Column 9*: Wormbase (WB) gene ID of the *C.elegans* ortholog. *Column 9*: Chromosomal location in *C. elegans*. *Column 10:* WB protein ID. *Column 11:* amino acid length of the longest protein encoded by the worm gene. *Column 12:* E score of BlastP worm versus human protein. *Column 13:* % length of the match. *Column 14:* result of the GFP-fusion experiment. *Column 15:* Result of the RNAi screen.

4.3.1 Selection of Candidate Genes

The selection of the genes to be analyzed included several steps. The primary criterion was the degree of homology of the human protein in nematodes. To be a suitable candidate, the human gene and its ortholog in *C. elegans* had to be highly similar. Thus, all human protein sequences of chromosome 21 (http://chr21.molgen.mpg.de/HSA21db.html) have been compared, using blastp program[19], to the whole protein catalogue of *C. elegans* (www.wormbase.com). The best *C. elegans* matches have been retrieved and "reverse blasted" against the whole human protein catalog (based on Ensembl). Only the best reciprocal matches in both organisms were selected and 89 orthologous gene of human chromosome 21 could be identified. The retrieved orthologs have then been screened for the highest conservation between human and nematodes as well as for their degree of conservation across mouse (*Mus musculus*), fruitfly (*Drosophila melanogaster*) and yeast (*Saccharomyces cerevisae*). An example of multiple alignments is shown in Figure 4-15 for one of the candidates (TMEM50B), where the protein sequences of the above cited organisms have been compared using ClustalW program. In addition, we prioritized genes with restricted expression patterns in mouse development (another running project in the laboratory on whole mounts *in situ* hybridization in mouse embryos (Gitton *et al.*, 2002)) and took into account the expression profiles obtained from the Ts65Dn analysis (described above). Finally, nine orthologs of Chromosome 21 in *C. elegans* have been selected, all of them being little or not functionally characterized so far (Table 4-2).

[19] Use the BLAST algorithm to compare an amino acid query sequence against a protein sequence database.

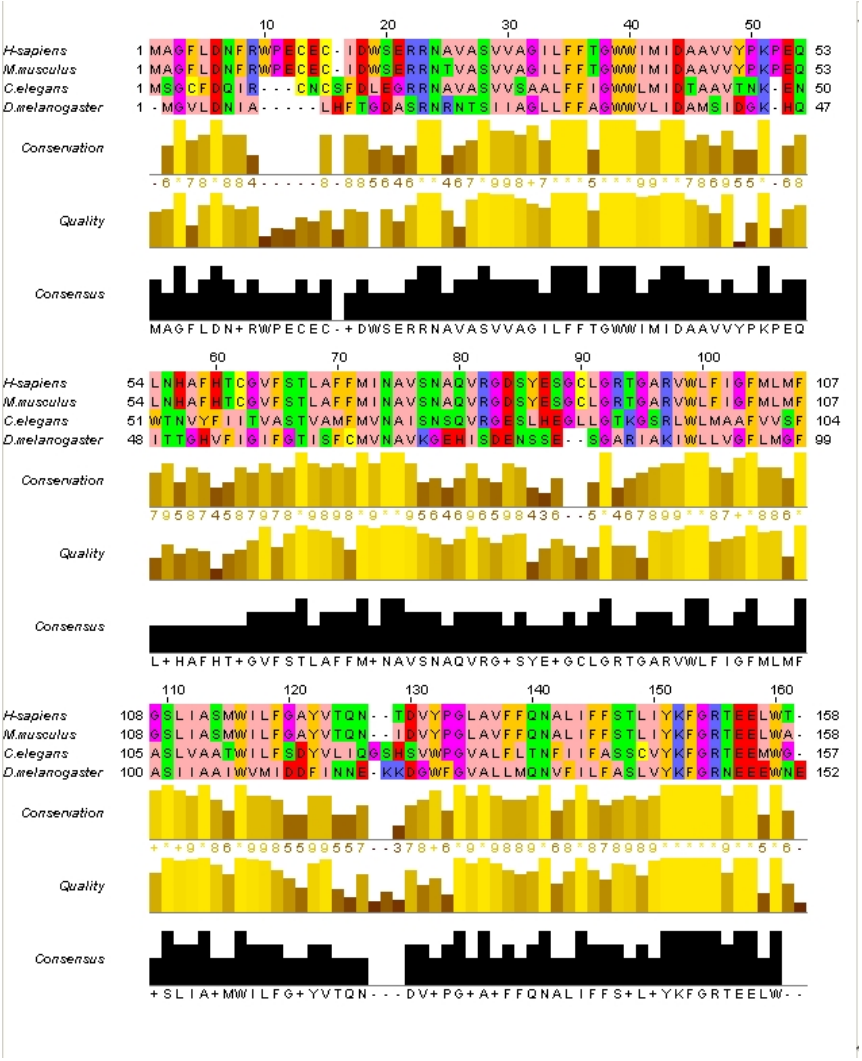


Figure 4-15: Multiple sequence alignment of TMEM50B and its orthologs. The multiple alignment was performed using clustalW program and the alignment editing was made with jalview 2.08b (Clamp *et al.*, 2004). Following references were used for the protein sequences: ENSP000000371390 for Homo sapiens, ENSMUSP00000023686 for Mus musculus, CG15012-PB for Drosophila melanogaster, WP:CE21753 for Caenorhabditis elegans. The residues are colored according to their physicochemical properties as follows: Aliphatic/hydrophobic residues are in *pink*; aromatic residues in *orange*; positive and negative residues in *blue* an *red* respectively; in *green* are hydrophilic residues; conformationally special residues are in *violet* and cysteine is in *yellow*. The alignment conservation annotation: This is an automatically calculated quantitative alignment annotation which measures the number of conserved physico-chemical properties conserved for each column of the alignment (Livingstone and Barton, 1993). Identities score highest, and the next most conserved group contains substitutions to amino acids lying in the same physico-chemical class. This means that the most conserved columns in each group have the most intense colors, and the least conserved are the palest. The alignment quality annotation is an ad-hoc measure of the likelihood of observing the mutations (if any) in a particular column of the alignment. More precisely, the quality score is inversely proportional to the average cost of all pairs of mutations observed in a particular column of the alignment - a high alignment quality score for a column would suggest that there are no mutations, or most mutations observed are favorable**.**

4.3.2 Gene Expression Localization by GFP fusion

In vivo expression was monitored via GFP (Green Fluorescent Protein) fusion constructs in order to determine when and where a particular gene is expressed. The strategy used to generate constructs containing the candidate gene fused to GFP and under the control of its own promoter is described in Figure 4-16. Briefly, primers were designed to amplify a genomic region including the full open reading frame of the gene of interest and a 5' upstream sequence up to the 3' end of the previous gene (or of at least 1kb). This ensures in most cases that the gene of interest is under the control of the endogenous promoter. Genomic DNA from wild type worms (N2), cosmids or yeast artificial chromosomes was used as template. The purified PCR product was cloned into an intermediate vector (Topo XL, Appendix Figure 7-5, p.161). The fragment was digested and inserted unidirectionnally into the L3781 vector (see Material and Methods and Appendix Figure 7-6, p.161) that contains a GFP coding sequence. The GFP protein was thus fused at the C-terminus of the candidate protein. All generated construct were partially sequenced and digested by restriction enzymes to verify their identity (list of all sequencing primers is given in Appendix Table 7-7, p.157). Transgenic animals were generated by co-injection of the DNA construct with a marker into the syncytial germ line (that is present in the distal gonad arm) of *lin-15* defective worm strains. Herein a genomic fragment of the *lin-15* gene was used as co-injection marker, which is probably the most commonly used rescuing marker. The loss of *lin-15* function results in a multivulva (Muv) phenotype that can only be scored in adults. The injected DNA concatemerizes, through both non-homologous end joining and through homologous recombination, generating a large, possibly circular DNA molecule. This new molecule usually does not integrate to one of the six chromosomes, but is rather maintained as an extrachromosomal array. Although the injected DNAs do not contain any obvious centromeric sequence or any clear origin of replication, the extrachromosomal array will be replicated and segregated at each cell division, behaving like a new minichromosome. This ability might reflect the fact that nematode chromosomes are holocentric[20]. The extrachromosomal arrays are transmitted with a lower fidelity than the endogenous chromosomes. A

[20] They do not possess a defined centromere, but rather kinetochores attached along each chromosome.

transmission rate of the transgene of 50% is considered as good. Transgenic animal lines could be generated successfully for five HSA21 orthologs (C33C12.9; Y74C10AL.2; C05D10.3, ZC373.1 and Y105E8A.1). Details of the protocols used herein are described in the Material and Methods section as well as a list well of all the generated worm lines and plasmids (Appendix Table 7-5 and Table 7-6, p.156). The expression of the remaining five selected genes could not be monitored either because the PCR fragment could not be amplified, or the cloning into the appropriate vector failed or no transmitted line could be obtained after injection.

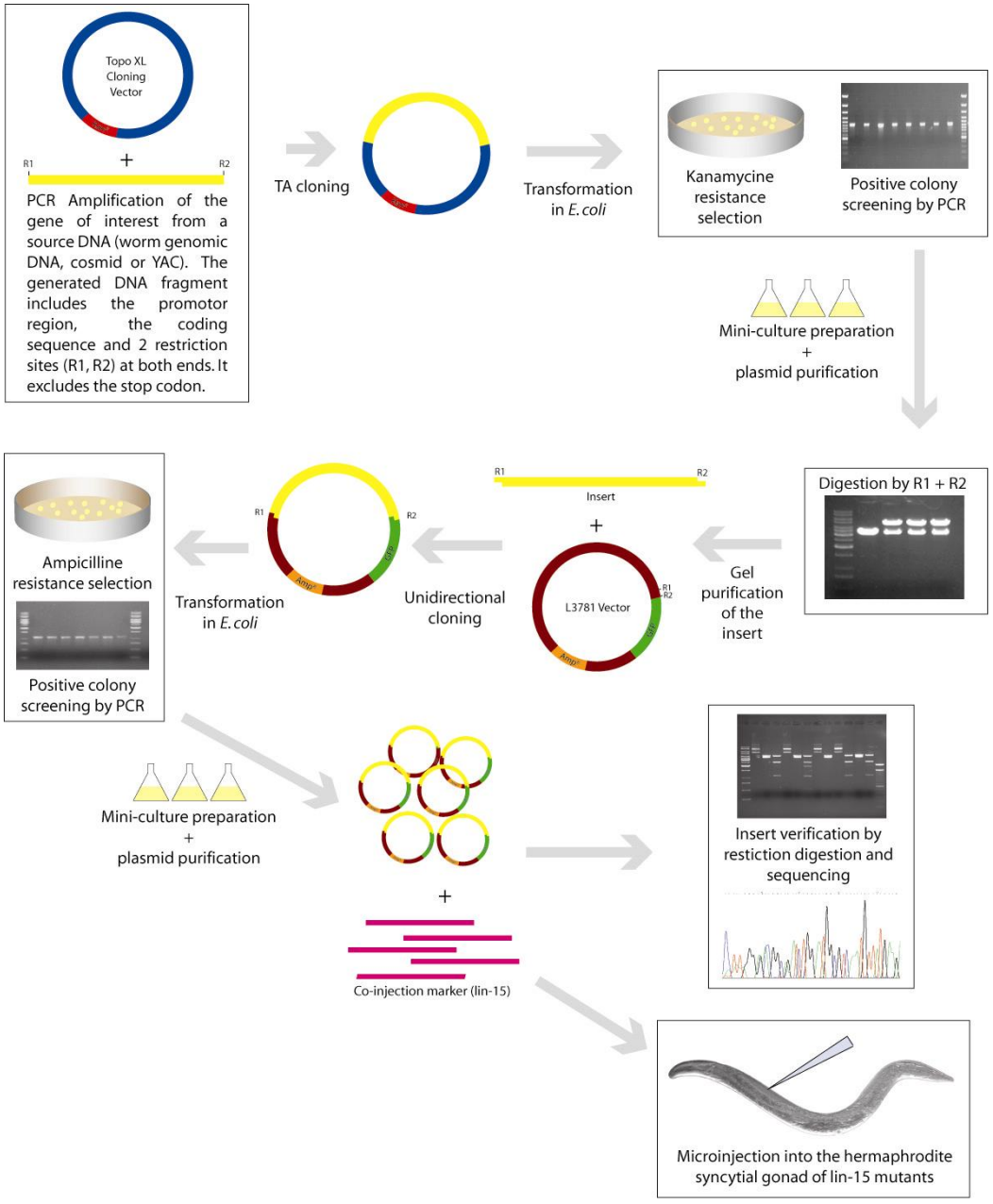


Figure 4-16: GFP fusion cloning strategy

All worm pictures in this section are lateral views of hermaphrodite worms and if note stated elsewise they are oriented such that anterior (head), posterior, tail, dorsal and ventral are left, right, top and bottom on the image.

Expression of the ortholog of CBS (ZC373.1)

In human, the protein encoded by CBS (565 Amino Acids (AA)) is involved in the transsulfuration pathway. It catalizes the first step of this pathway, which converts homocysteine to cystathionine. CBS deficiency can cause homocystinuria that affects many organs and tissues, including the eyes and the skeletal, vascular and central nervous systems. In *C. elegans*, the ortholog of CBS (ZC373.1) encodes two splice variants (ZC373.1.1 and ZC373.1.2) composed of 10 exons each and differing only by their 3' untranslated region. Both encode the same 704 amino acid related protein. It shares similar domains to human protein but has yet not been functionally characterized in the worm. To characterize the endogenous expression of ZC373.1 in *C. elegans* we cloned a 3.1 kb genomic region in L3781, including the complete coding sequence of ZC373.1, as well as a 1.8 kb region upstream the start codon (Figure 4-17.A). Three lines could be retrieved and analyzed (Appendix Table 7-5, p.156), all of them showing similar expression patterns. We observed for ZC373.1, a strong cytoplasmic expression in intestine, in the head and bodywall muscles from the larval state until adulthood (Figure 4-17.B-G). In 2006, McKay and colleagues reported an equivalent expression pattern for this gene as well as expression in hypodermis using a GFP-endogenous promotor reporter assay (McKay *et al.*, 2003).

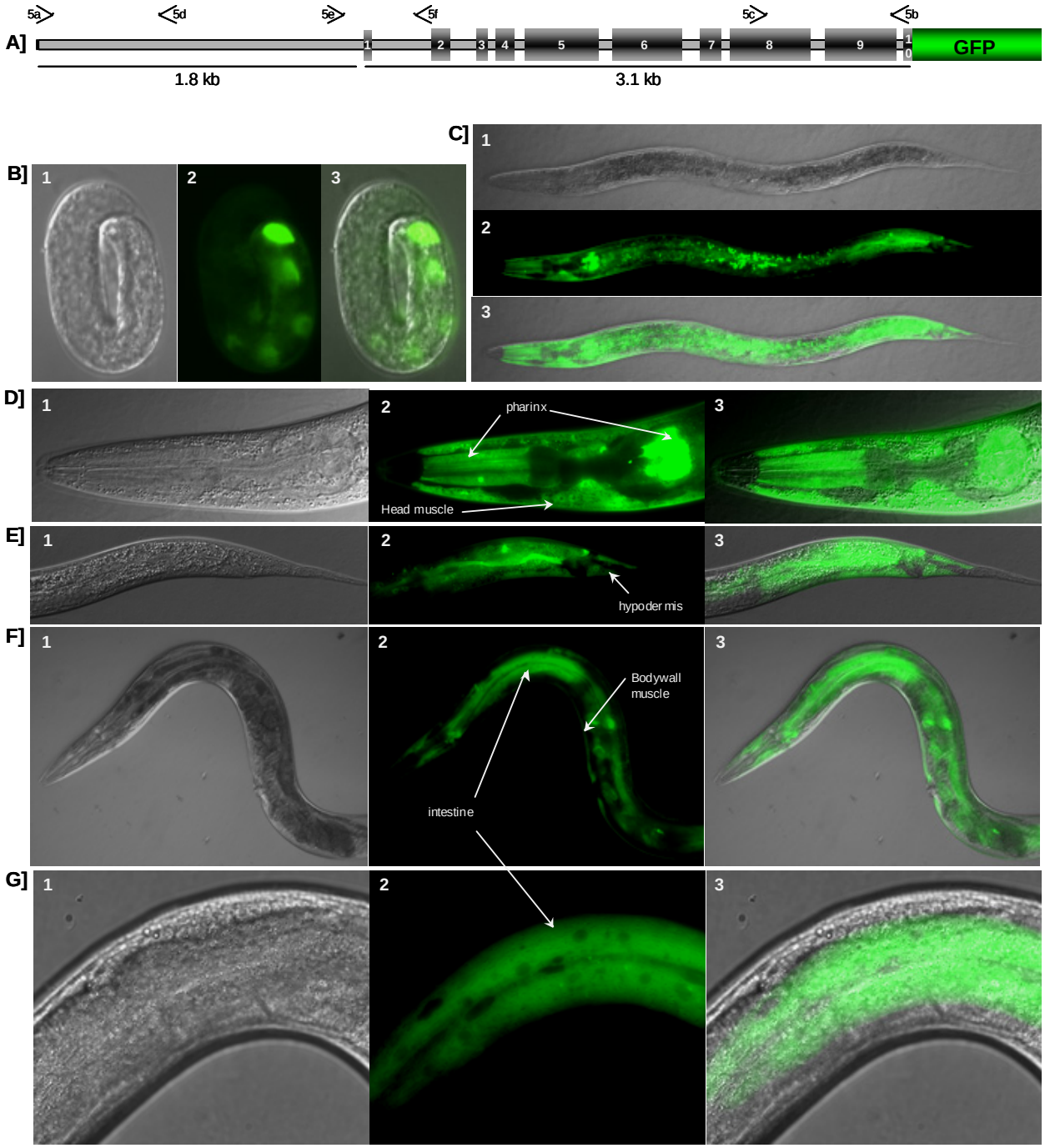


Figure 4-17: Endogenous expression of the human CBS ortholog. A) 5´ to 3´ Gfp fusion construct (ZC373.1::GFP) including the coding sequence of ZC373.1 (3.1 kb) and 1.8 kb upstream the start codon. The arrows with numbers indicate the primers used to generate the PCR product (5a & 5b) and those used for sequence identity verification (all). For each serie **1)** shows the normal DIC image, **2)** the GFP only image and **3)** the superposed picture of DIC and GFP. **B)** Embryonal expression in intestinal cells at 63x magnification. **C)** Expression in an L2 animal at 10x magnification. **D)** Expression in pharinx and head muscle of an L2 animal at 63x magnification. **E)** Expression in intestine and hypodermis of the tail of an L2worm at 40x magnification. **F)** Expression in the adult worm at 10x magnification. **G)** Cytoplasmic expression in intestine of an adult at 40x magnification.

Expression of the ortholog of *HEMK2* (C33C12.9)

The protein (214 AA) encoded by the HEMK2 gene in human putatively belongs to the methyltransferase superfamily. The worm ortholog of HEMK2 referenced as C33C12.9 encodes a unique splice form composed of 4 exons producing a protein of 234 AA. To date, excepted for a predicted DNA methyltransferase domain nothing is known about the function of protein encoded by this gene neither in human nor in worm. To generate the GFP fusion construct we cloned a 5.7 kb genomic region that included the complete coding sequence of C33C12.9, as well as a 3.1 kb region upstream the start codon (Figure 4-18.A) in L3781 vector. Four lines were generated (Appendix Table 7-5, p.156) and they showed similar expression patterns. Interestingly this reporter gene has a highly specific pattern of gene expression localized only in some parts of the nervous system from L1 larval stage until adulthood (Figure 4-18.B-K). The neuronal expression observed herein was restricted to the ventral nerve cord, in some head and tail neurons and along the commissures. In the worm, most of the nerve cords run as longitudinal process bundles along the main body axis. However, individual neuron processes can make transitions between these longitudinal process trajectories. These transitions, which run circumferentially (dorso-ventrally) between the longitudinal nerves are called commissures.

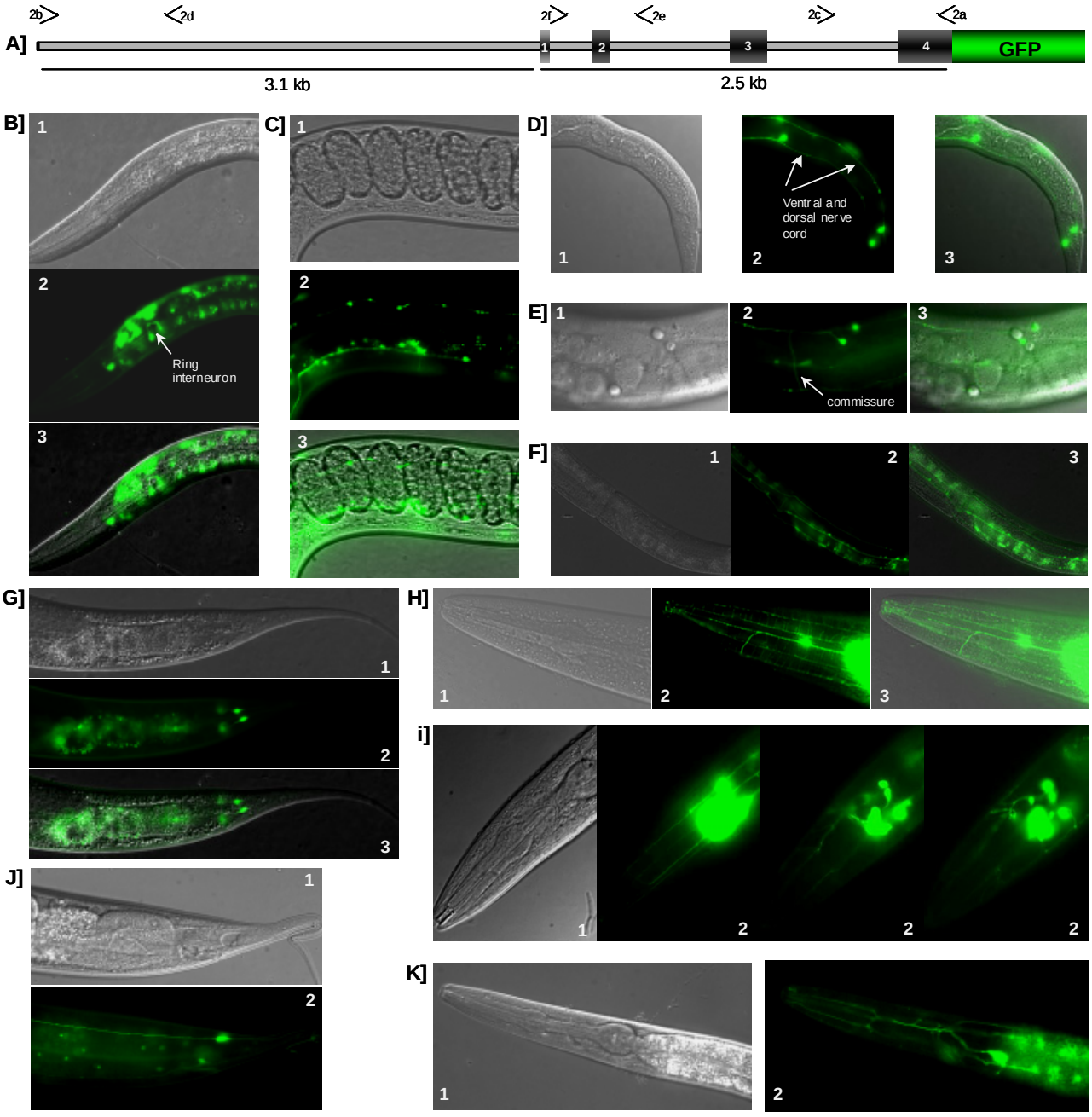


Figure 4-18 : Endogenous expression of the human HEMK2 ortholog. A) 5´ to 3´ Gfp fusion construct (C33C12.9::GFP) including the coding sequence of C33C12.9 (2.5 kb) and 3.1 kb upstream the start codon. Numbered arrows indicate the primers used to generate the PCR construct (2a & 2b) and those used for sequence identity verification (all). For each serie **1)** shows the normal DIC image, **2)** the GFP only image and **3)** the superposed picture of DIC and GFP. **B)** Expression in head neurons (nerve ring) of an L1 animal at 40X magnification. **C)** Expression in a young adult at 40X. **D)** Expression in the nerve chord of an L1 individual at 63x. **E)** Expression in commissures of an adult at 63X. **F)** Expresion in young adult at 40X. **G)** Expression in the tail of a young adult lying on the belly at 40x (dorsal view). **H)** Neuronal expression in the head of an adult at 63X. I) Different layers of neuronal expression in the head of an L4 animal at 63X. **J)** Neuronal expression in the tail of an adult at 40X. K) Neuronal expression in the head of an L4 worm at 40X.

Expression of the ortholog of TMEM50B (Y74C10Al.2)

The protein encoded by this gene in human (158 AA) as well as in the worm is predicted to be a transmenbrane protein of unknown function. To date, nothing is known about the function of this protein. The worm ortholog of TMEM50B, Y74C10AL.2, encodes two splice variants (Y74C10AL.2.1 and Y74C10AL.2.1.2) composed of 3 exons each and differing only by their 3' untranslated region. Both encode the same 157 amino acids long protein. To investigate the endogenous expression of this protein, we cloned a 4.3 kb genomic region that included the complete coding sequence of Y74C10AL.2, as well as a 1.4 kb upstream region (Figure 4-19.A) in L3781 vector. Three worm lines expressing the GFP fusion protein could be successfully retrieved (Appendix Table 7-5, p.156), with similar expression patterns. The highly specific and restricted expression of Y74C10Al.2 was detected in neuronal precursor cells of embryos (Figure 4-19.B-C). Until the L4 stage this expression remained localized to specific neurons and to the hypodermis (Figure 4-19.B-F). From L4 stage till adulthood a very strong and consistent cytoplasmic expression was detected in seam cells and in hypodermal nuclei, whereas the neuronal expression seemed to decrease (Figure 4-19.G-K). Only few neuronal cells showed a rather weak expression. The external surface of *C. elegans* is formed by hypodermis, which establishes the basic body form of the animal, acts in nutrient storage, secretes the cuticle and takes up apoptotic cell bodies by phagocytosis. Many of the hypodermal cells form multinucleate syncytia that are formed by cell fusions during development. Hypodermal cells can be described in two general categories:

1- Hypodermis, which is formed by a main body syncytium, hyp7, and other hypodermal cells in the head and the tail, numbered from hyp 1 through hyp 11

2- Specialized epithelial cells: Several types of supporting cells which might be of either hypodermal or glial character are included here. Some may act as potential guides for axon outgrowth, protectors of neuronal sensory receptors, or as linker cells to hold various portions of the hypodermis together. Many produce specialized adaptations of the cuticle, to permit the creation of holes or ridges in the cuticle. Some do not produce any cuticle products. These cells fall into 3 subcategories:

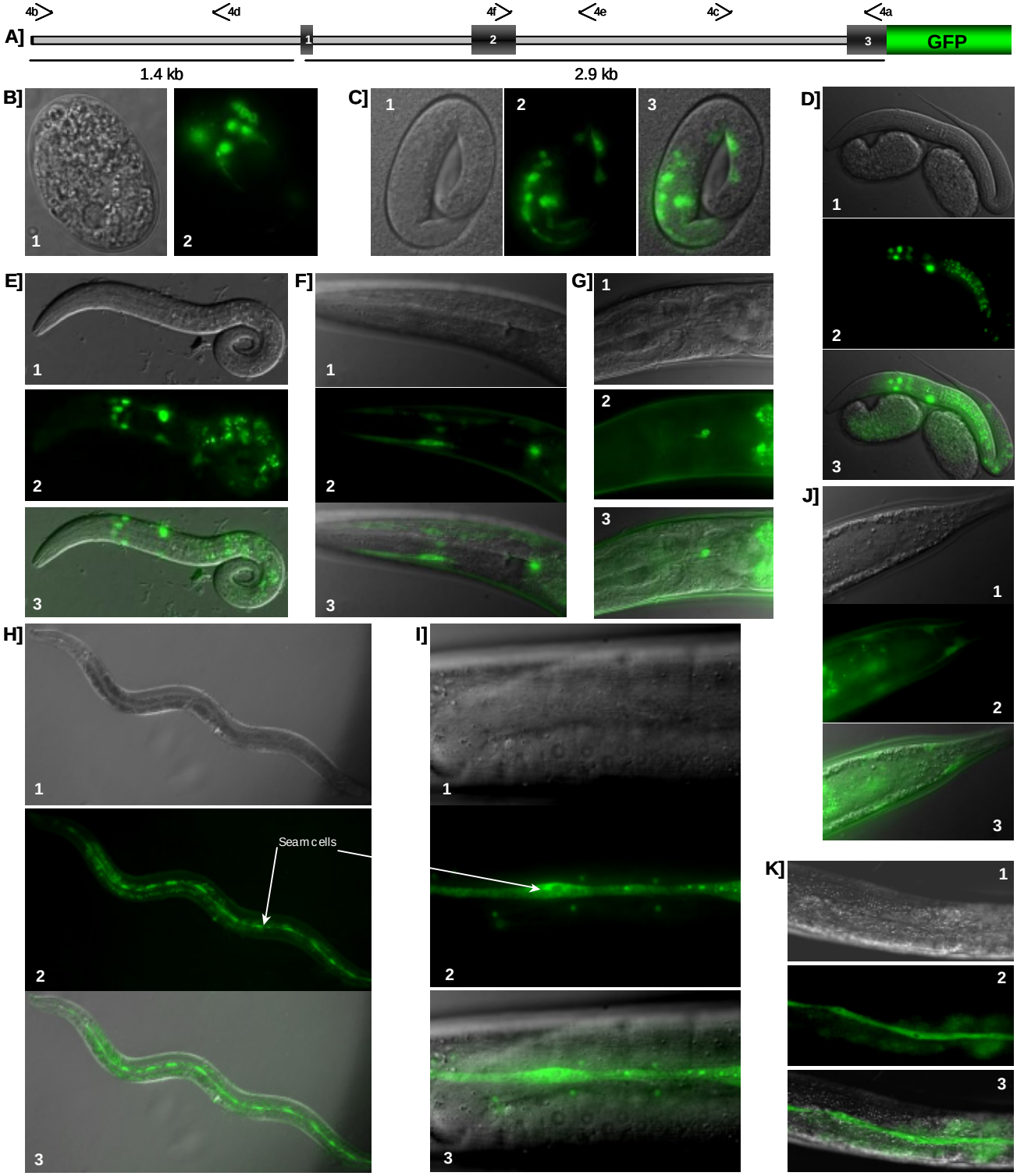
a- Seam cells, which lie along the apical midline of the hypodermis, at the extreme left and right sides between nose and tail. During postembryonic development, they can act as stem cells to produce neurons of lumbar ganglia, tail spike neurons and phasmidial support cells, and sensory bristles along the sides. They are also responsible for the production of the cuticular alae in L1 stage, dauer larvae, and adults.

b- Interfacial epithelial cells: These include cells that are located at the junctions where hypodermis meets another type of tissue.

c- Other epithelial cells.

The expression pattern observed herein suggests that Y74C10AL.2 is temporally and spatially highly controlled.

Figure 4-19 (Next page): Endogenous expression of the human TMEM50B ortholog. A) 5‘ to 3‘ GFP fusion construct (Y74C10AL.2::GFP) including the coding sequence of Y74C10AL.2 (2.9 kb) and 1.4 kb upstream the start codon. Numbered arrows stand for primer name and sense that were used for the generation of the PCR product (4a & 4b) and those used for the sequence identity verification (all). For each serie **1)** shows the normal DIC image, **2)** the GFP only and **3)** the superposed picture of DIC and GFP. **B)** and **C)** Expression in early and late embryo at 63X. **D)** and **E)** Neuronal expression in L1 worms at 63X and 40X magnification respectively. F) Expression in the head of an L3 individual at 63X. **G)** Neuronal expression in the head of an adult at 63X. **H)** Expression in seam cells of an adult worm at 10X. **I)** Seam cells in adult at 63X **J)** Expression in the tail of an adult (dorsal view) at 63X. **K)** Seam cells in adult at 40X.


A]
4b
4d
4f
4e
4c
4a
1
2
3
GFP
1.4 kb
2.9 kb
B]
C]
D]
E]
F]
G]
H]
I]
J]
K]
Seam cells

Expression of the ortholog of SH3BGR (Y105E8A.1)

In human, SH3BGR encodes a 239-amino acid protein characterized by the presence of a proline-rich region containing the consensus sequence for an SH3-binding domain and by an acidic carboxy-terminal region containing a glutamic acid-rich domain predicted to conform to a coiled-coil structure. The presence of two functional domains involved in protein-protein interactions suggested that SH3BGR might be part of a multimeric complex in human. The protein is differentially expressed in heart and skeletal muscle (Gitton *et al.*, 2002; Scartezzini *et al.*, 1997). The worm ortholog, Y105E8A.1, encode a unique splice form composed of 4 exons. The resulting protein is 132 amino acids long. To investigate the endogenous expression of this protein, a 6.9 kb genomic region was cloned in L3781, including the complete coding sequence of Y105E8A.1 and the 3.9 kb upstream region (Figure 4-20.A). Nine lines were generated (Appendix Table 7-5, p.156). McKay and colleagues reported a larval expression in intestine, posterior cells, seam cells, nervous system, ventral nerve cord, head neurons and unidentified cells in head (McKay *et al.*, 2003). In contrast we detected a strong cytoplasmic expression from L1 on to adulthood (Figure 4-20.B-I) in bodywall muscle, pharinx, head muscles, sex muscles, anal sphincter and depressor muscles and in some adults expression was also observed in head neurons. No expression was detected in embryos.

Though most muscle contractions are generated by nerve transmission, there are three rhythmic behavior cycles in *C. elegans* that are dependent on periodical contraction of certain muscle groups with recurrent intracellular Ca++ transients rather than excitation by neuronal transmission. These are the pharyngeal pumping behavior of the pharyngeal muscle, the gonadal sheath contractions, and the defecation cycle involving body wall (somatic) muscles near the head, posterior (somatic) body wall muscles and enteric muscles (i.e., anal depressor, sphincter and stomato-intestinal muscles). There are eight distinct muscle divisions in the pharynx, each containing one to three muscle cells (Figure 4-21.A). Contraction of pharyngeal muscle cells serves to open the lumen allowing the uptake of food. To a certain extend the pharyngeal muscle is comparable to the heart muscle in humans.

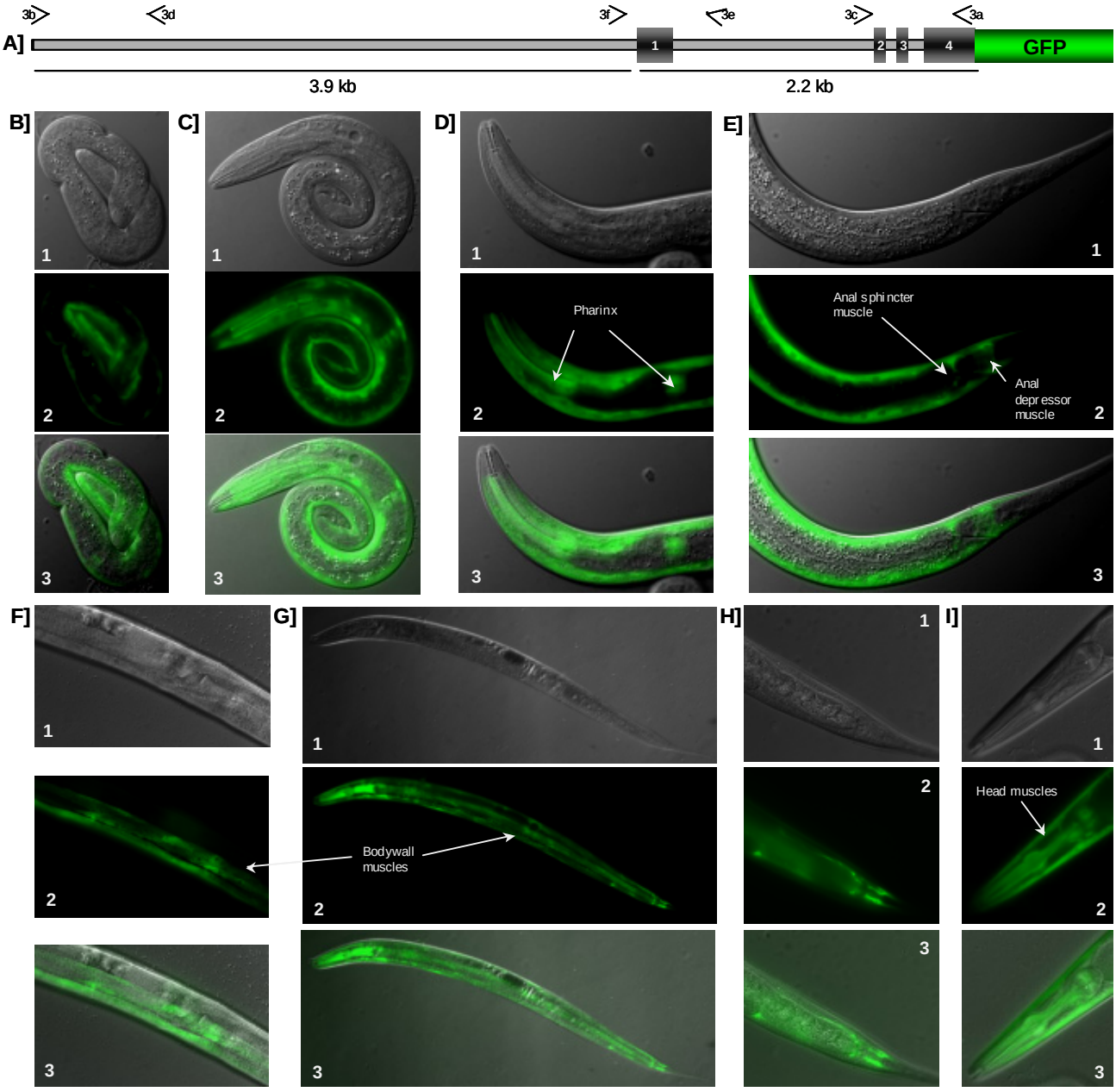


Figure 4-20: Endogenous expression of the human SH3BGR ortholog. A) 5´ to 3´ GFP fusion construct (Y105E8A.1::GFP) including the coding sequence of Y105E8A.1 (2.2 kb) and 3.9 kb upstream the start codon. The arrows with numbers indicate the primers used to generate the PCR product (3a & 3b) and those used for sequence identity verification (all). For each serie **1)** shows the normal DIC image, **2)** the GFP only image and **3)** the superposed picture of DIC and GFP. **B)** Cytoplasmic expression in bodywall muscle in 3 fold embryo at 63X magnification. **C)** Expression in bodywall and head muscle and in the pharinx of an L1 worm at 63X .**D)** Cytoplasmic expression in bodywall and head muscle and in the pharinx of an L2 worm at 63X. **E)** expression in the anal sphincter and anal depressor muscle of an L2 animal at 63X. **F)** Bodywall muscle expression, ventral view of L3 worm at 40X. **G)**. L3 worm at 10x magnification (ventral view) **H)** Anal depressor and sphincter muscle expression (ventral view) in L3 individual at 40X. **I)** Expression in the head muscle of an adult at 40X.

The anal sphincter muscle circles the intestine at its junction with the rectum (Figure 4-21.B). It is dilated prior to the enteric muscle contractions during defecation and contracts near simultaneously with the other enteric muscles, possibly acting to further squeeze the posterior intestine. The anal depressor (Figure 4-21.B) muscle is a large, single sarcomere, H-shaped muscle in hermaphrodites that runs vertically between the dorsal wall of the rectum and the dorsal hypodermis. This muscle lifts the roof of the rectum when it contracts, hence allowing the rectum to fill during initial stages of defecation. Later during defecation, it relaxes and the contents of the rectum are expelled. The body wall muscle of *C. elegans*, as in all other nematodes, is obliquely striated. Although the filaments themselves are oriented parallel to the longitudinal plane of the muscle cell, adjacent structural units. It is composed of 95 rhomboid-shaped body wall muscle cells arranged as in four longitudinal bundles. The expression pattern observed in muscles of *C.elegans* is similar to the expression seen in humans. Like the bodywall muscles of the worm, skeletal muscles are striated muscles used to facilitate movement, by applying force to bones and joints via contraction. In humans, the cardiac muscle, like other muscles, can contract, but it can also conduct electricity, like nerves. This allows the rhythmical pumping behavior, which is also seen in the worm's pharyngeal muscle. Altogether these observations suggest that along to protein domain conservation of SH3BGR in worms and humans, the molecular function might also be highly conserved.

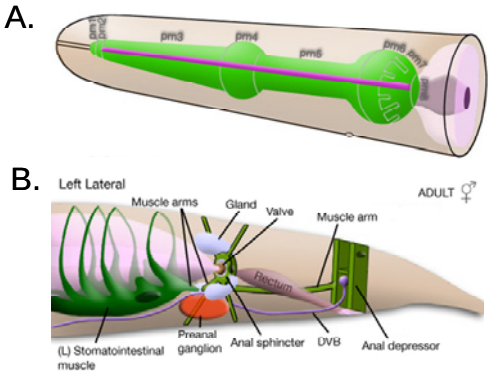


Figure 4-21: Muscles in *C. elegans.* A) Paryngeal muscle cells (pm) in head of an adult hermaphrodite worm **B)** Muscles in theTail of an adult hermaphrodite worm. Pictures were taken from www.wormatlas.org.

Expression of the ortholog of ABCG1 (C05D10.3)

In human, the protein encoded by ABCG1 gene is a member of the superfamily of ATP-binding cassette (ABC) transporters and the longest variant is composed of 825 AA. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member of the White subfamily, which is involved in macrophage cholesterol and phospholipids transport, and may regulate cellular lipid homeostasis in other cell types. Several alternative splice variants have been identified. The worm orthologous gene, C05D10.3, encode a unique splice form composed of 11 exons. The resulting protein is 598 amino acids long. To investigated the endogenous expression of this protein, a 4.1 kb genomic region was cloned in L3781, including the complete coding sequence of C05D10.3 and the 1.4 kb upstream region (Figure 4-22.A). Five lines were generated (Appendix Table 7-5, p.156), but no expression could be detected in these lines at any stage (Figure 4-22.B). McKay and colleagues reported in wormbase a larval expression in the intestine that could not be confirmed here (McKay *et al.*, 2003). They used a GFP-promotor reporter assay that included the same promotor region as in our assay (the 1.4 kb upstream region of the start codon.) Furthermore, the correct identity of our construct was assessed by sequencing and the gene could neither be detected by PCR on worm mRNA population (see Chapter 3.5.3). Nonetheless it is possible that 1.4 kb upstream region didn't include the promotor region, that the expression of this gene was too weak to be detected by this method, or that it is restricted to a very specific time window that has been missed in the analysis.

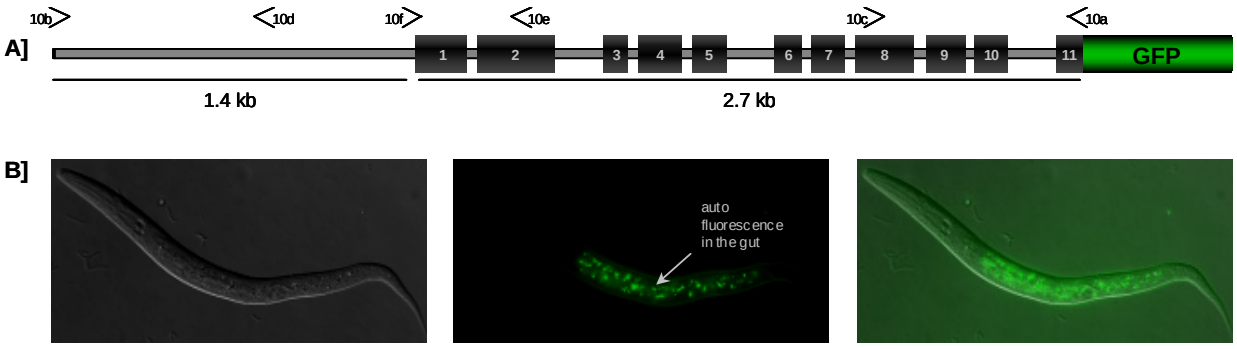


Figure 4-22: Endogenous expression of the human ABCG1 ortholog. A) 5´ to 3´ Gfp fusion construct (C05D10.3::GFP) including the coding sequence of C05D10.3 (2.7 kb) and 1.4 kb upstream the start codon. The arrows with numbers indicate the primers used to generate the PCR product (10a & 10b) and those used for sequence identity verification (all). **B)** Lateral view of an L1 hermaphrodite worm at 40X magnification. No obvious expression is detected. For each serie **1)** shows the normal DIC image, **2)** the GFP only image and **3)** the superposed picture of DIC and GFP.

4.3.3 Gene Silencing by RNAi

RNA interference or RNAi is a well-established technique to knock down target gene expression post-transcriptionally through degradation of the target messenger RNA (Fire *et al.*, 1998). RNAi has become a very useful tool in functional genomics. In *C. elegans*, three methods are used to apply RNAi: By feeding, soaking or injecting. Herein we used feeding, since it is the most convenient method. The worm is well suited for RNAi experiments as it can simply be fed with bacteria expressing antisense RNA against the gene of interest. The antisense approach to gene silencing involves incorporating into an organism an RNA sequence complementary to mRNA transcribed from a target gene. The antisense RNA and sense mRNA hybridizes and are degraded by the RISC protein complex. The presence of dsRNA duplex led to what has now been recognized as an RNA interference (RNAi) effect. Using *C. elegans* we aimed to silence the expression of the nine selected candidate genes by RNAi techniques in order to track for possible consequences on the phenotype of the developing worm. Several groups have by now performed large-scale RNAi screens in *C. elegans* and their results will are being compared to ours herein (Fernandez *et al.*, 2005; Fraser *et al.*, 2000; Kamath *et al.*, 2003; Kim *et al.*, 2005; Rual *et al.*, 2004; Simmer *et al.*, 2003; Sonnichsen *et al.*, 2005). The strategy used to generate the *E. coli* strains expressing siRNAs[21] directed against targets is described in Figure 4-23: RNAi experimental strategy.. Briefly gene specific PCR fragments were amplified from *C.elegans* cDNA, cloned in an appropriate vector (L4440) for RNAi (see Material & Methods and Appendix Figure 7-7, p.162) and transformed into HT115 *E. coli* strain. The wild type worm strain (N2) and two RNAi sensitive strains (VH525, RRF-3) were than fed with the generated *E. coli* strains. For seven candidate genes a strain expressing siRNAs could be generated. For two genes (C05D10.3, Y105E8)

[21] siRNA stands for small interfering RNA.

the PCR fragment could not be successfully generated. As controls we used a *E. coli* strain transformed with the empty L4440 vector that should not give any phenotype, and a strain expressing siRNAs against Unc-22 that alters the movements of the worms (Uncoordinated) when it is silenced. To control whether the genes of interests were effectively silenced, we used, when available, the corresponding GFP-fusion worm lines (see chapter 3.5.2) to monitor the reduction in fluorescence (Figure 4-24). For C33C12.9 and C47E12.7 we used additional L4440 constructs from the RNAi library of the Ahringer lab (Kamath and Ahringer, 2003) (Appendix Table 7-9). They produced the same phenotype in *C. elegans* as our constructs. We induced all siRNAi expressing *E. coli* strains directly on plates as well as in liquid culture, the latter enabling a stronger inducing effect.

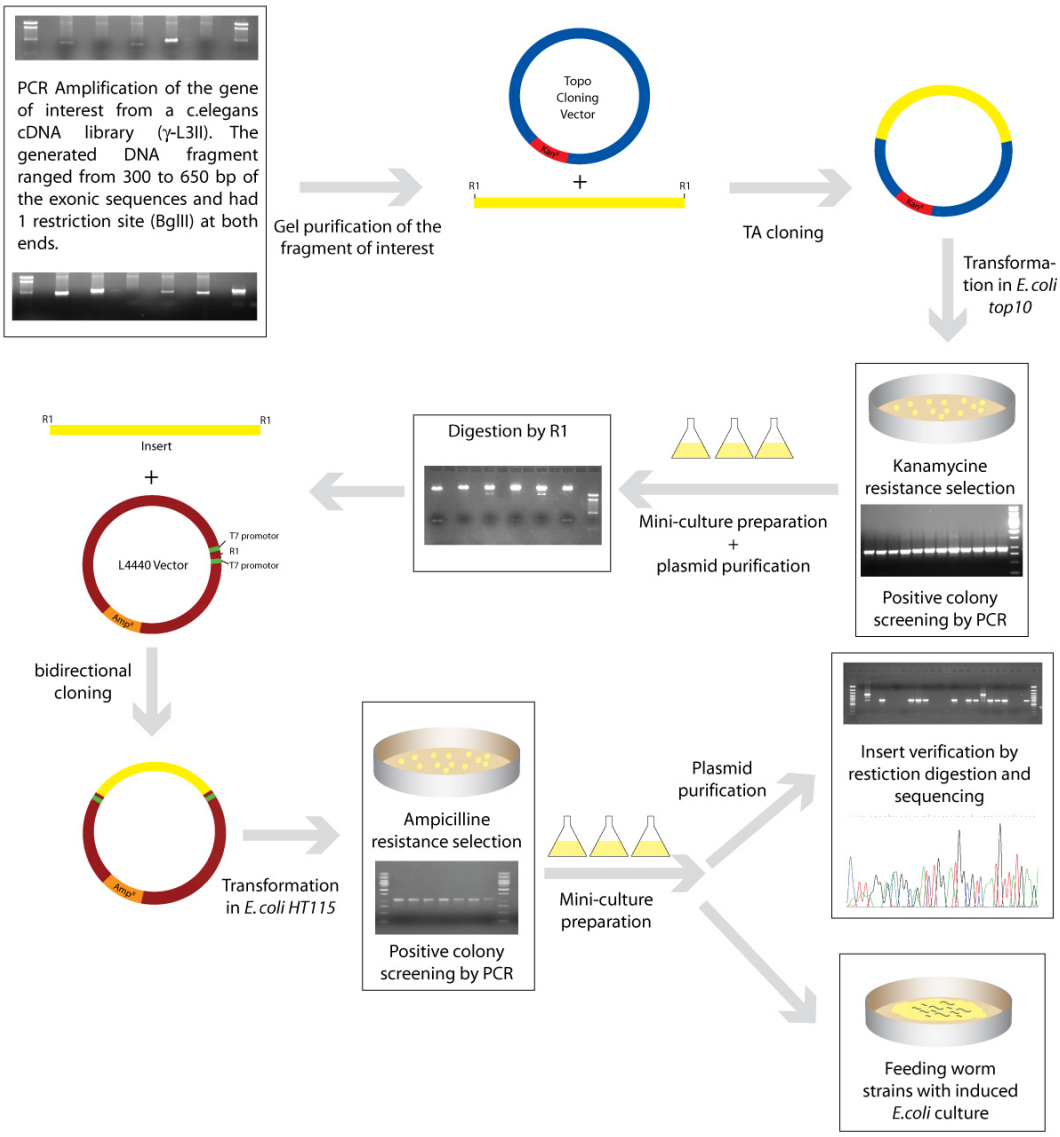


Figure 4-23: RNAi experimental strategy.

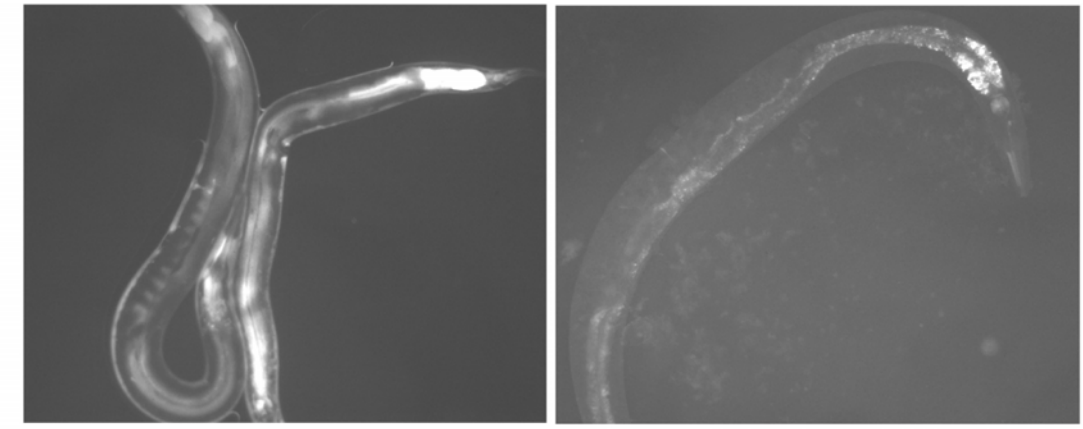


Figure 4-24: Silencing ZC373.1 expression by RNAi. A) ZC373.1::GFP fusion worm line grown on *E.coli* carrying the empty L4440 vector. **B)** ZC373.1::GFP fusion worm line grown on *E.coli* carrying the siRNA expressing L4440::ZC373.1 vector.

Overall, the gain of knowledge of the results of the RNAi experiments (Table 4-3) remained limited. Only one gene (C47E12.7) produced a visible phenotype when silenced. C47E12.7 is the ortholog of the human protein encoded by KIAA0179, which has to date no known function. The three wormlines in which this gene was silenced showed a potential migration defect (Mig), slow growth (Gro) and an abnormal larval development at an early stage (Lva) where the larvae arrest development during the L1 stage. In recent, RNAi screens several groups observed analogous phenotypic consequences (Kamath and Ahringer, 2003; Simmer *et al.*, 2003). Additionally Fernandez *et al.* observed embryonic lethality and Rual and colleagues observed maternal sterility (Fernandez *et al.*, 2005; Rual *et al.*, 2004). The latter phenotypes were also scored but not observed by Sonnichen B *et al.* (Sonnichsen *et al.*, 2005). These findings suggest that C47E12.7 is involved in biological processes controlling the embryonic and larval development, the growth and the reproduction. Furthermore, protein domain recognition identifies it as a nucleolar protein. The six remaining genes didn't give an observable phenotype when silenced. However, a significant reduction in fluorescence was observed for ZC373.1 in the intestine and for Y74C10AL.2 in the seam cells when the corresponding GFP fusion worm lines were fed with the siRNA expressing strains. This suggests that the silencing method worked effectively for these genes. For ZC373.1, Sonnichsen *et al.* (Sonnichsen *et al.*, 2005) observed a late larval arrest

where nearly all progeny arrest before L4 larval stage, that could not be confirmed herein. In contrast we didn't detect a substantial reduction in fluorescence in neurons for C33C12.9, suggesting that the siRNA directed against this gene could not reach their target. However, because feeding RNAi is thought to work inefficiently in *C. elegans* neurons, neuronal-specific genes are unlikely to be recovered in our screen. About C24H12.5, Kamath *et al.* (Kamath *et al.*, 2003) report a larval lethal phenotype as well as an abnormal locomotion. This could not be assessed herein and was neither confirmed by other reports (Rual *et al.*, 2004; Sonnichsen *et al.*, 2005) scoring the same phenotypes. F28B3.1 is also subject to contradictory reports as Simmer *et al.* observed a 10% death rate of embryos during embryonic development in the hypersensitive *rrf-3* strain, but other groups did not confirm this (Fraser *et al.*, 2000; Kamath *et al.*, 2003; Sonnichsen *et al.*, 2005). Finally, for Y116A8C.36 the phenotypic examination of worms did not reveal an obvious abnormality compared with wild type (Fernandez *et al.*, 2005; Kamath *et al.*, 2003; Rual *et al.*, 2004; Sonnichsen *et al.*, 2005), as in our screen.

RNAi targets | Worm strains	L4440-ZC373.1 (CBS)	L4440-C24H12.5 (DONSON)	L4440-Y116A8C.36 (ITSN)	L4440-Y74C10AI.2 (C21ORF4)	L4440-F28B3.1 (KIAA0184)	L4440-C47E12.7 (KIAA0179)	L4440-C33C12.9 (C21ORF127)
N2	WT	WT	WT	WT	WT	Mig/ Gro/ Lva	WT
VH525	WT	WT	WT	WT	WT	Mig/ Gro/ Lva	WT
RRF-3	WT	WT	WT	WT	WT	Mig/ Gro/ Lva	WT

Table 4-3: RNAi results. For each RNAi target are shown the results in the three tested worm lines N2, VH525 and RRF-3. The name of the corresponding human ortholog is in brackets. Results obtained by the wild type (WT), migration defect (Mig), growth defect (Gro), Larva arrest (Lva).

Overall the information gained from our RNAi screen remained little. Many of our tested genes seemed either to be resistant to RNAi or the phenotypic consequences were not detectable in our system. To further tackle information about our candidates, *C.elegans* knock-out strains for Y74C10AI.2 and C33C12.9

have been ordered from a Japanese knock-out consortium (http://shigen.lab.nig.ac.jp/c.elegans).