Chapter 4

Genetic variation and the expression of the OR repertoire.

I have demonstrated that high-depth RNAseq is a suitable technology to study the transcriptome of the olfactory system, and that the expression of the majority of the receptor repertoire can be detected, albeit at low levels. The amount of sequencing data devoted to the receptor genes can be increased by separating the OSNs from the other cell types found in the tissue, but the increase is only modest (2.5 fold). These experiments revealed that the profile obtained from the whole tissue is virtually identical to that from sorted OSNs, indicating that despite the lower amount of data obtained from WOM samples, the information is accurate and reproducible. The advantage of analysing the transcriptome of the WOM is that wild-type animals can be used, without the need for any transgenes. Therefore, I decided to continue to profile WOM samples. Also, since the OR genes are more divergent than VRs, these can be accurately detected by unique counts, and thus I focused my subsequent analyses on the OR repertoire only.

The data presented so far has revealed a characteristic expression distribution for the OR repertoire; a few genes are highly abundant and expression values then drop quickly. Remarkably, each of several animals from the same genetic background contains a stereotyped proportion of each receptor type, suggesting that OR gene choice is very tightly regulated. But how rigid is this pattern of expression and what are the factors contributing to such a stable state?

4.1 Gender has little effect on OR gene expression.

The C57BL/6J (hereinafter B6) WOM transcriptome data presented in Chapter 2 was obtained from adult male and female samples. Among the behavioural responses elicited through olfactory signals, many are clearly distinct between adult male and female mice, including sexual conduct[157, 159, 323], aggressive responses to intruders[156] and parental care[128], but the mechanisms that ensure such differentiated responses have not been fully elucidated in mammals[324]. It is conceivable that the receptors involved in the detection of the chemicals mediating these dimorphic behaviours might be differentially expressed between sexes. To assess this, I performed a differential expression analysis. Surprisingly, the transcriptomes showed a striking similarity between males and females (Figure 4.1). At the whole genome level, only 32 genes were significantly DE (FDR $< 5\%$). Among these are those expected to be different by sex, such as genes from the Y chromosome or the X-inactive specific transcript, *Xist*. Except for these, the fold-changes observed were small, with only a mean 1.56 difference between the sexes, for both genes expressed higher in males and in females. The expression estimates for each OR gene between males and females were also remarkably similar, with only 0.7% of the repertoire significantly DE (9 OR genes; Figure 4.2). This suggests that the observed behavioural dimorphism is not achieved by a differential ability to detect certain olfactory cues.

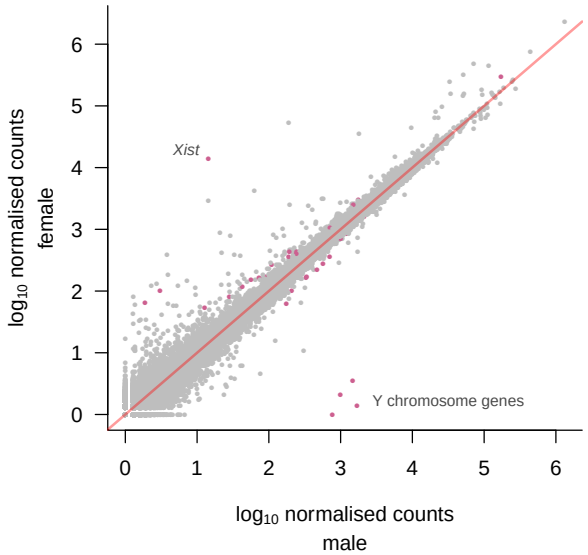

Figure 4.1 – Transcriptome of the WOM of males and females. Scatter plot of the mean normalised counts for the transcriptome of the WOM in males and females. The red line indicates the 1:1 diagonal. Most genes lie close to the diagonal, suggesting that the transcriptomes are nearly equivalent. The 32 significantly differentially expressed genes are highlighted in pink; among these are genes known to be expressed only in one sex, such as *Xist* or genes from the Y chromosome.

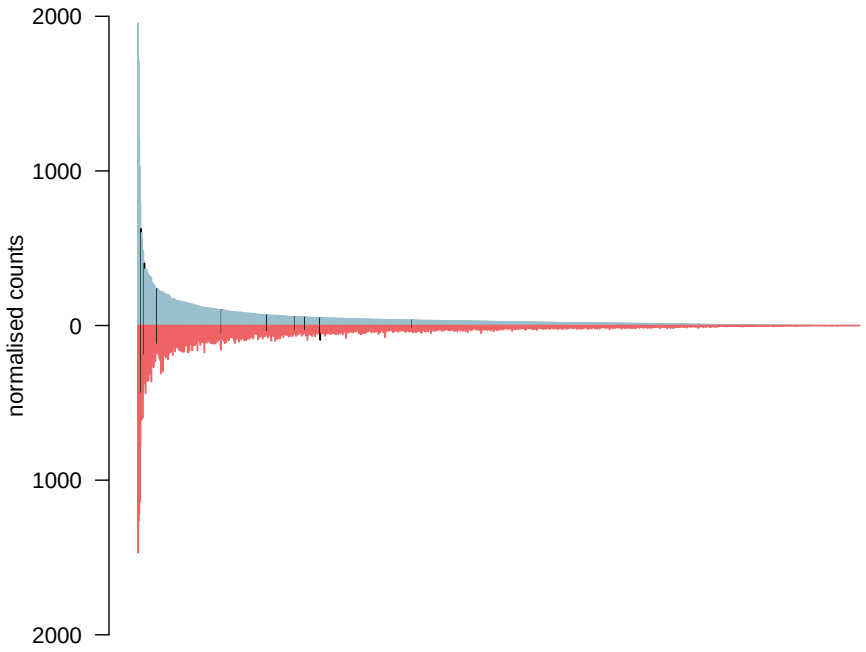

Figure 4.2 – OR expression in males and females. Barplot of the normalised expression of the OR genes in males (blue) versus females (red). Counts have been normalised for both depth of sequencing and gene length. Genes are ordered by decreasing male expression, and the corresponding values in females are presented as a mirror image. The repertoires are extremely similar. Significantly differentially expressed genes are highlighted in black.

4.2 Some OSN types are more abundant than others.

The observed distribution of expression values for the OR genes is conserved between individuals and is not significantly affected by sex. Such an expression profile may be explained by two scenarios: either 1) OR genes with high expression values are expressed in a larger population of neurones than those with low expression values; and/or 2) OR genes are consistently expressed at different levels per cell. To differentiate between these possibilities, I first compared the RNAseq expression values with cell counts in the MOE, for a subset of OR genes. These data were obtained by counting fluorescent OSNs from transgenic mice (21 days of age) carrying tagged receptors[325]. Since all these transgenic lines are in a mixed 129×B6 genetic background similar to the OMP-GFP line, I compared to the RNAseq data from the OMP-GFP WOM (25 days of age) from Chapter 3. The correlation with the RNAseq normalised counts was very strong and highly significant (rho = 0.94, p-value = 0.00047; Figure 4.3A). Thus, the expression values obtained through RNAseq are a reflection of the number of cells expressing a particular receptor gene within the MOE.

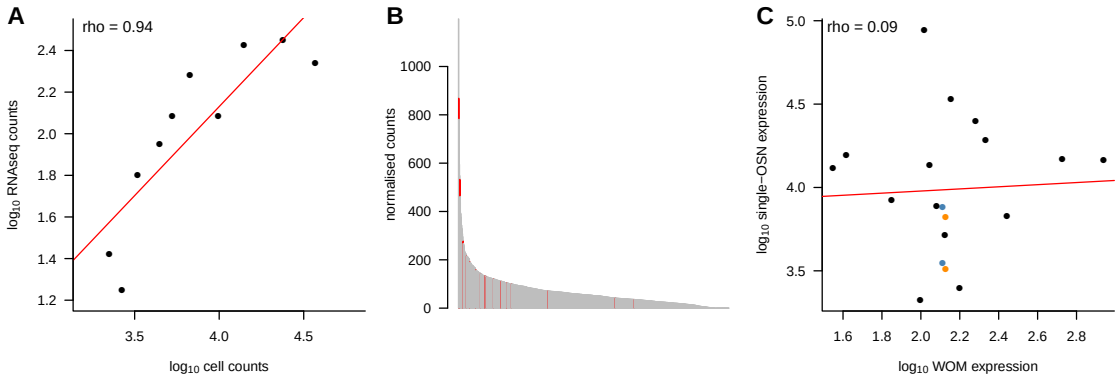

Figure 4.3 – RNAseq expression correlates with neurone number A) The expression levels for a subset of OR genes are plotted against their count number in the MOE. In red is the regression line and the Spearman correlation coefficient is on the top left corner. The high correlation indicates that RNAseq expression levels reflect the number of cells expressing a particular OR gene in the MOE. **B)** Barplot of the expression of the OR repertoire in the OMP-GFP WOM data, in decreasing abundance order. Those receptors that were expressed in the 19 single OSNs sequenced are highlighted in red. These cluster towards the highly expressed genes which suggests that the MOE has higher number of OSNs expressing these receptor genes, which results in a higher probability of sampling them. **C)** The RNA expression in the OMP-GFP WOM data is plotted against the expression levels of abundant OR genes in single OSNs. The two single cells expressing *Olfr55* are in orange and the two expressing *Olfr728* are in blue. In red is the regression line. There is no correlation between the two, suggesting that the cumulative expression seen at the population level is not influenced by the OR abundance per OSN.

These data indicate that the first scenario might be the major contributor to the observed expression profile for the receptor repertoire. If this is true, one would expect that the OR genes expressed in randomly sampled neurones from the MOE would be enriched for those expressed at high levels. Such a random sampling of OSNs was performed when producing the single-cell RNAseq data presented in Chapter 3. Indeed, from the 17 different OR genes expressed at high levels in the 19 single OSNs, 14 are within the top quartile of the distribution and this enrichment is highly significant (p-value = 9.23e-07, hypergeometric test; Figure 4.3B). Thus, there is a strong bias towards selecting receptors with high RNAseq expression estimates within the WOM. To test if there is also a contribution from varying levels of OR expression per OSN, I compared the expression levels in the WOM RNAseq data against the expression of each OR in the individual OSNs; to account for variability, I normalised to five OSN marker genes (*Omp*, *Gnal*, *Adcy3*, *Ano2* and *Cnga2*) that have been shown to be stably expressed[304]. While the OR gene expression levels did vary within single cells (mean = 15,353.57 and SD = 19,483.62 normalised counts), they did not correlate with the corresponding expression levels in the WOM (rho = 0.09 p-value = 0.71; Figure 4.3C). Together, these data suggest that while there is variance in the OR expression levels between different OSNs, this does not correlate to their cumulative abundance in the overall population. Instead, the expression levels obtained from WOM samples reflect the number of cells expressing a particular OR gene. Therefore, RNAseq is an accurate proxy for quantifying

the diversity of OSN types found within the MOE, such that high RNAseq expression equates to high number of OSNs.

4.3 OR expression differs between mouse strains.

The relative proportion of each OSN type is stable between genetically identical animals irrespective of sex (Figure 4.2). To investigate whether this OSN distribution is a feature of all mice, I analysed the WOM transcriptome of a different laboratory strain, 129S5SvEv (referred to as 129; Tables B.1 and B.2 in Appendix B). The 129 genome has 4.4 million single nucleotide polymorphisms (SNPs) and 810 thousand small indels (1-100 bp long) compared to B6[321], of which 13,484 SNPs and 1,936 indels are found within OR gene transcripts. All these variants can have an impact on the mapping of the RNAseq data, especially since ORs are more variable than the average gene[271]. To assess this, I mapped the data to the B6 reference genome, or to a pseudo-129 genome that contains all the high quality SNPs and short indels reported for this strain[321]. Despite the great majority of the repertoire having similar expression estimates in both cases (85.6% have changes of less than one normalised count), a few receptor genes show a significant change in the number of sequencing fragments that map to them (Figure 4.4). In most of these cases (84.4%), the pseudo-129 genome allows a greater proportion of reads to be mapped. Thus, I utilised the data mapped to the pseudo-129 genome for further analyses.

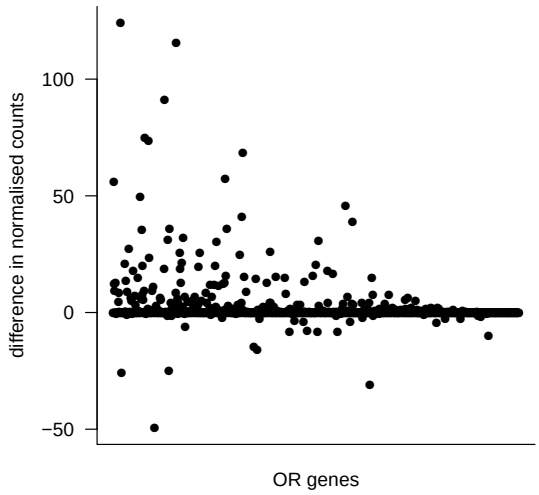

Figure 4.4 – Effect of imputing variation on OR expression estimates. Plot of the difference of the mean expression values for OR genes, as obtained by mapping to a pseudo-129 genome versus mapping to the B6 reference. The genes are ordered by their decreasing mean expression value in the calculations using the pseudo-129 genome.

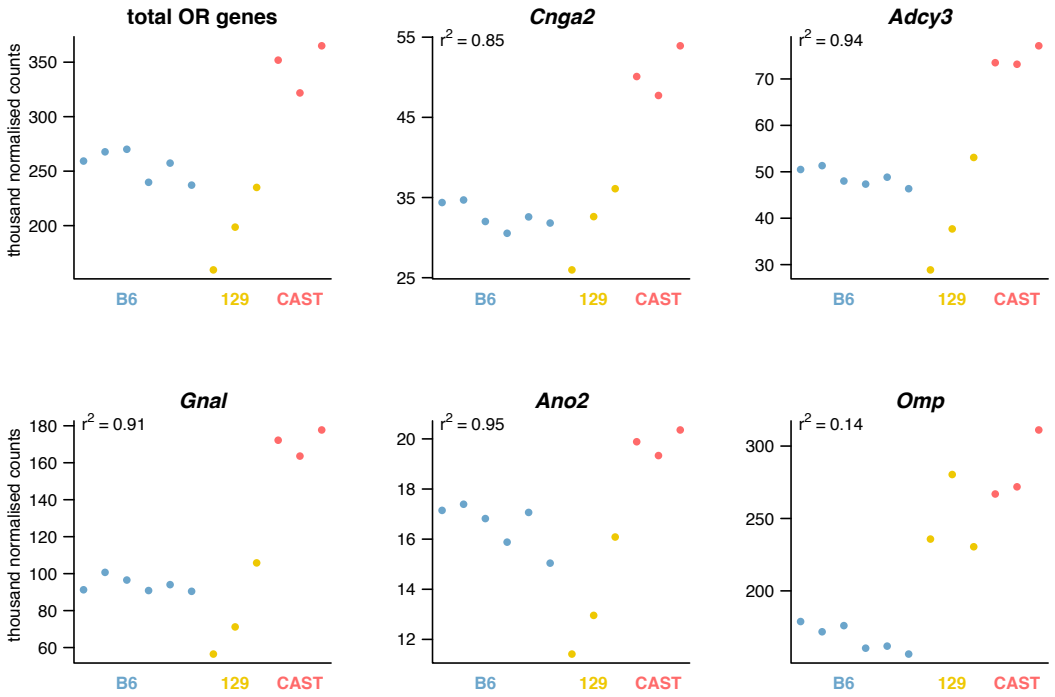

Figure 4.5 – Normalisation for OSN number. In the first panel the total normalised counts for OR genes are plotted for each WOM sample from different strains; CAST data will be presented below. Then, the normalised counts are plotted for a set of marker genes that are known to be stably expressed in OSNs only: *Cnga2*, *Adcy3*, *Gnal*, *Ano2* and *Omp*. The Pearson's correlation value with the total counts is indicated. All marker genes except for *Omp* are highly correlated to the total number of counts in OR genes, suggesting that the observed differences are the product of different proportions of OSNs in the WOM samples.

To be able to compare the relative distribution of OR gene expression in mature OSNs across different strains, it is necessary to normalise to account for any differences in the total number of neurones present in WOM samples. For this, I used a method proposed by Khan et al.[304]. This approach uses marker genes known to be stably expressed in mature OSNs only, to estimate the proportion of WOM RNA contributed by the OSNs. As can be observed in Figure 4.5, the total number of normalised counts across the receptor repertoire varies considerably between strains, even after normalising for differences in depth of sequencing. Inspection of genes only expressed in OSNs (*Cnga2*, *Adcy3*, *Gnal*, *Ano2* and *Omp*) revealed that most of them follow the same pattern of expression and are highly correlated with the total OR counts (Figure 4.5). Therefore, the observed differences are most likely due to different proportions of OSNs in the WOM samples. In this particular case, *Omp* had a very poor correlation value ($r^2 = 0.14$, p-value = 0.22, Pearson) and thus was not used in the normalisation process. To normalise for OSN number, size factors were obtained based on the geometric mean of the OSN markers, and these were used to scale the OR genes normalised counts.

To evaluate how similar are the OR expression distributions between the animals

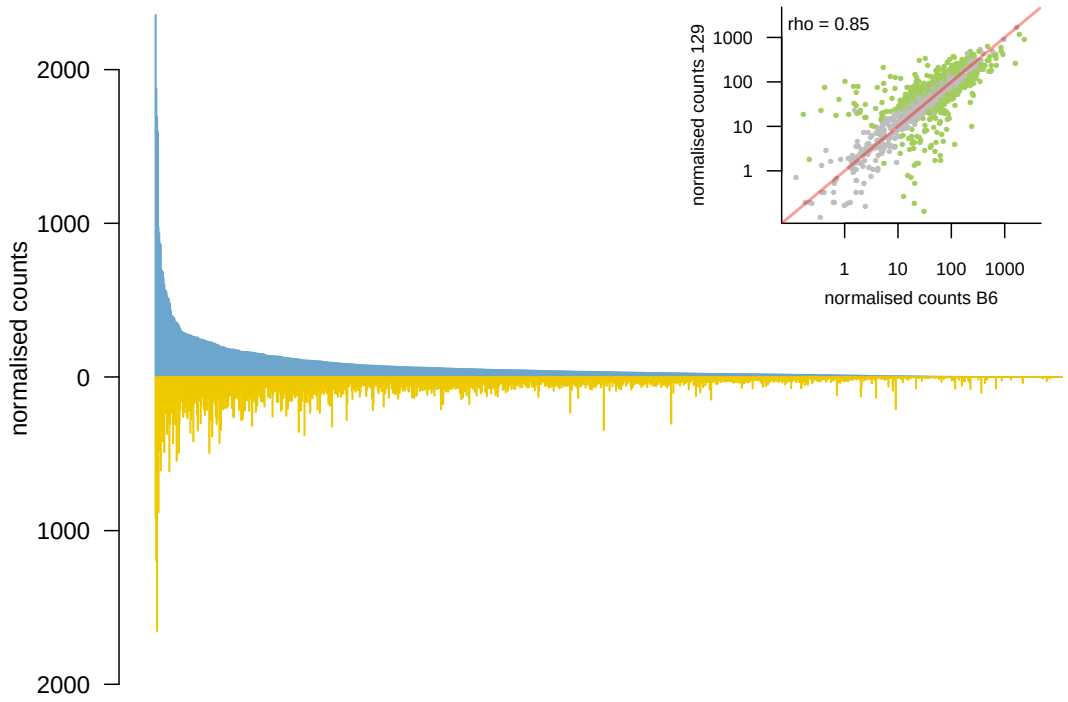

Figure 4.6 – OR expression in B6 and 129. Barplot of the OR normalised expression values in B6 (blue) with the corresponding 129 values (yellow) as a mirror image. As an inset is the scatter plot of the same expression levels; the red line is the 1:1 diagonal and the Spearman correlation coefficient is indicated. Significantly differentially expressed genes (FDR < 5%) are highlighted in green. Many OR genes have differences, albeit small, in their expression levels.

from these two genetic backgrounds, I performed a differential expression analysis on all OR genes. From these, 462 were significantly DE (FDR < 5%), which represents 37% of the whole repertoire (Figure 4.6). However, 45.9% of the DE genes had a difference lower than 2-fold, implying consistent but relatively small changes in expression. Concordant with this amount of DE genes, the correlation for OR gene expression was 0.85 (p-value < 2.2e-16, Spearman), that contrasts with the very high correlation of 0.96 (p-value < 2.2e-16, Spearman) across the whole transcriptome.

To determine whether genetic diversity influences the variance in OR gene expression, I repeated this experiment using CAST/EiJ, a wild-derived strain from the *Mus musculus castaneus* subspecies (henceforth referred to as CAST; Tables B.1 and B.2 in Appendix B). This strain has 17.6 million SNPs and 2.7 million short indels relative to B6; of these, 45,688 SNPs and 6,303 indels are found within OR transcripts. After mapping to a pseudo-CAST genome, 634 OR genes were significantly DE (FDR < 5%) compared to B6, constituting 50.8% of the whole OR repertoire (Figure 4.7). The changes in expression for some OR genes were dramatic: 132 genes had differences of at least 8-fold and therefore the OR gene correlation between the two strains was only 0.73 (p-value <

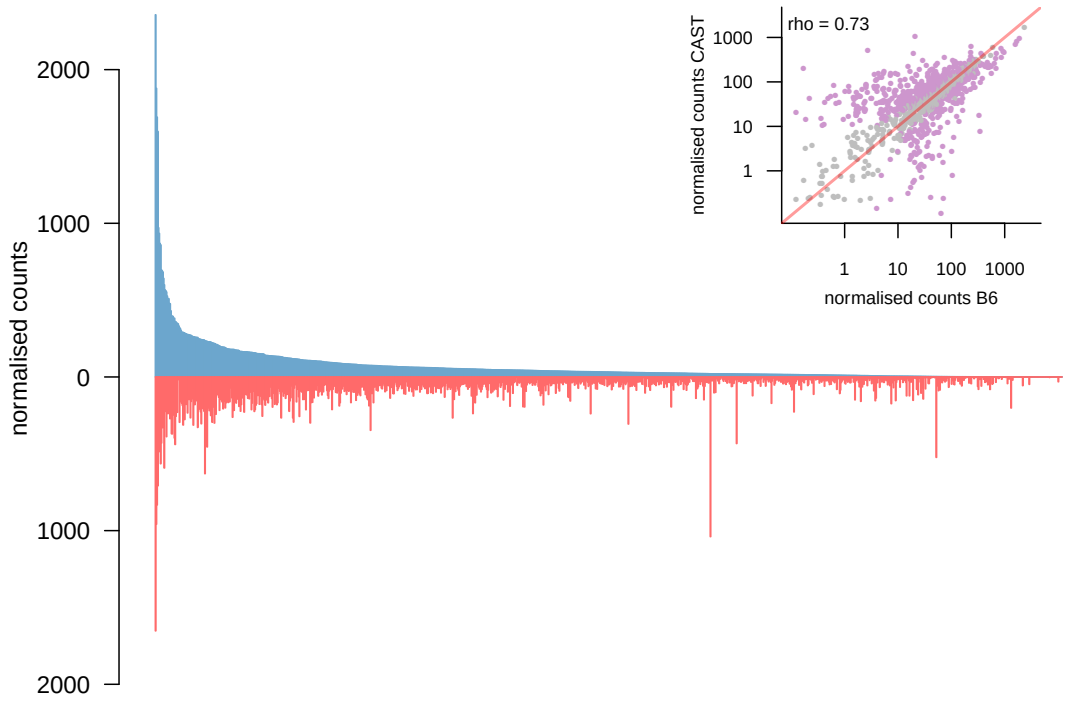

Figure 4.7 – OR expression in B6 and CAST. Barplot of the OR normalised expression values in B6 (blue) with the corresponding CAST values (red) as a mirror image. As an inset is the scatter plot of the same expression levels; the red line is the 1:1 diagonal and the Spearman correlation coefficient is indicated. Significantly differentially expressed genes (FDR < 5%) are highlighted in purple. Many OR genes have distinct expression levels and some are strikingly different.

2.2e-16, Spearman) while at the whole genome level the correlation remained high, at 0.96 (p-value < 2.2e-16, Spearman).

Taking all pairwise comparisons into account (including 129 vs CAST), 821 OR genes (65.7%) were DE between at least two strains (Figure 4.8A). Of these, 136 were DE in all three comparisons (Figure 4.8B), indicating a consistently different level of OR gene expression in each strain. All together, I have profiled the complete OR repertoire of three different strains of mice, with varying levels of genetic divergence among them. The data indicates that the OR expression levels are tightly controlled; a very stable distribution is conserved among animals of identical genetic background. However, the composition of the neuroepithelium of the WOM is remarkably diverse between animals with a different genetic makeup. Since the differences in expression levels indicate differences in the number of OSNs expressing particular OR genes, the MOE of each strain is a mosaic composed of varying proportions of each of the ~1,000 OSN types.

In order to investigate how the genetic background could be affecting OR expression levels, I mined the Mouse Genomes Project catalogue of SNPs and short indels for the 129 and CAST genomes[321]. Differentially expressed OR genes had greater amounts of

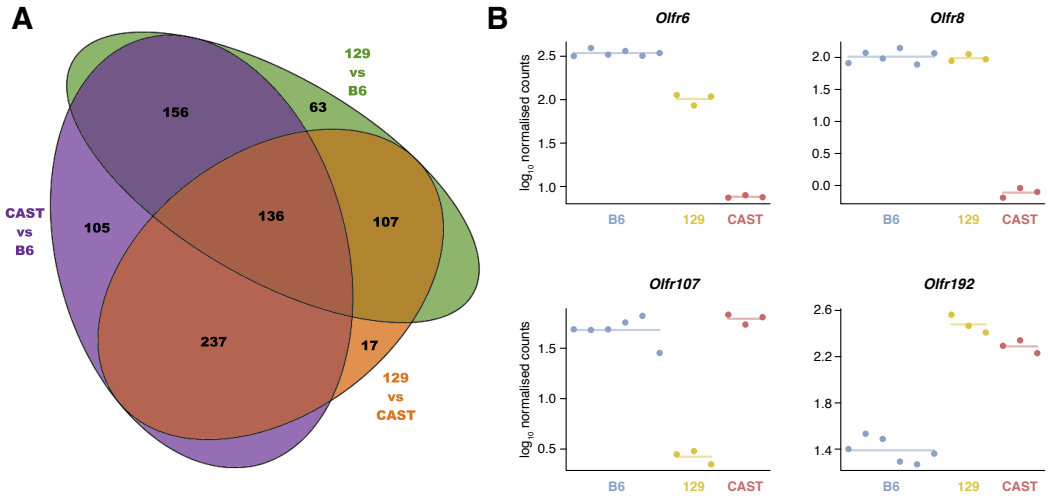

Figure 4.8 – Differential expression of the OR repertoire in three strains of mice. A) Proportional venn diagram indicating the number of OR genes that are significantly DE in the different pairwise combinations between the three strains. 136 receptor genes have different expression levels in all three strains. **B)** Examples of OR genes that are differentially expressed. Each point is a biological replicate and the horizontal line indicates the mean expression for each strain. In order, a receptor with differing expression values in each strain, one that is different in CAST only, in 129 alone or different in B6.

variation in their CDS, whole transcript or regions of 300bp or 1kb upstream of the TSS, in comparison to non-DE genes (Mann-Whitney, one tail), for both the 129 and CAST genomes (Figure 4.9). However, there was no relationship between the number of variable positions and the fold-change between the strains. The analysis of the putative promoters of many OR genes has shown that these have several OE-like and homeodomain binding sites[192–194], and the latter have been demonstrated to influence the frequency with which an OR gene is chosen for expression[188]. Therefore, it is likely that higher amounts of variation increase the probability of affecting some of these motifs, which would result in changes in the overall expression in the MOE.

Variation can also impact the open reading frame directly. For example, the B6 genome contains a premature stop codon in *Olfr421-ps1*, which results in a truncated protein of only 269 amino acids. On the other hand, the 129 and CAST genomes have a SNP that reverts the stop codon into a coding amino acid and, therefore, both produce a full-length protein (315 amino acids). OSNs that initially choose a pseudogene switch to another functional OR gene and progressively extinguish the expression of the non-functional receptor[207]. Consistent with this, the B6 animals have very low expression of *Olfr421-ps1*, while it is expressed more than ten fold higher in the 129 and CAST WOM (Figure 4.10).

Several examples exist on the influence of genetic variation on OR gene choice, such as the H-element which has been shown to elicit stronger effects the closer it is to the

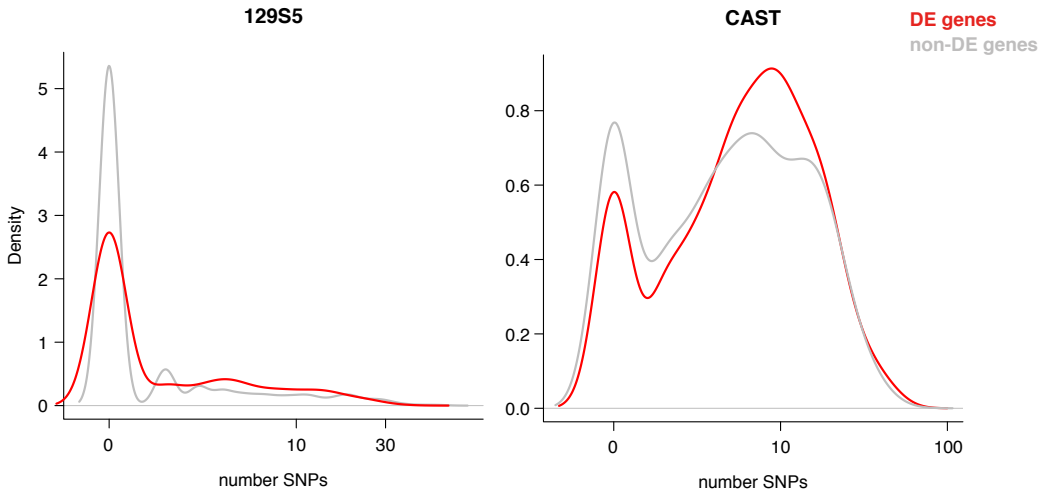

Figure 4.9 – Differentially expressed OR genes are more variable. Density distribution of the number of OR genes that have a given number of SNPs in the 1kb region upstream of the TSS. In grey are all those genes that are expressed at equivalent levels between B6 and 129 (left) or CAST (right); and in red, are the significantly differentially expressed genes (FDR $< 5\%$). There are less DE genes that have no SNPs and greater numbers of DE receptors have larger numbers of SNPs.

OR cluster it regulates[182]. The distance in the 129 genome is greater than in B6 animals[183] and, therefore, the expression levels of the proximal genes of the MOR28 cluster are lower in 129 than in B6. Enhancers, like the H-element and others, also contain transcription factor binding sites[188]. They can be located tens to hundreds of kilobases away from the OR genes they regulate. Commonly, they affect several OR genes

Figure 4.10 – Expression of a polymorphic pseudogene. *Olfr421-ps1* is an OR gene that in the B6 genome has a premature stop codon, but produces a full-length protein in the 129 and CAST animals. On the top is the gene model; the thicker box indicates the CDS and the rest is annotated UTRs. The blue line indicates the SNP that produces the STOP codon in B6. Below are representative examples of the raw sequencing data obtained from B6, 129 or CAST WOM. Each read is drawn in grey and the blue lines indicate the read spans exon-exon junctions; from the data it can be inferred that the gene model is lacking an additional 5' non-coding exon. Coloured lines within the reads indicate mismatches. The expression is many fold higher in the strains with a functional gene.

from the most proximal cluster[183, 186, 187]. Inspection of the pattern of differential expression in terms of the cluster organisation of OR genes revealed that, often, several adjacent genes are regulated in a concerted manner, with all genes expressed at higher or lower levels in a given strain. One such example is presented in Figure 4.11; a cluster of 8 OR genes, 6 of which are functional, consistently show lower expression in B6 than in 129 and CAST.

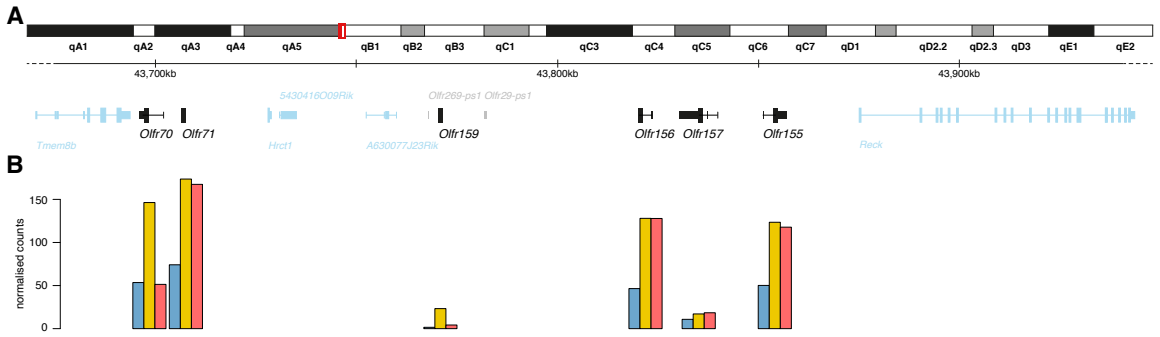

Figure 4.11 – Differentially expressed OR genes are clustered in the genome. A) Schematic representation of a cluster of OR genes. On the top is chromosome 4 and the region in the red box is expanded below. Each gene model is drawn with boxes for exons and connecting lines for introns. The height of the box indicates whether the sequence is protein-coding (taller) or not. A cluster of 8 OR genes spans 162.6kb; six of the genes are annotaed as functional (black) and two are pseudogenes (grey). Non-OR genes are shown in blue. **B)** The normalised expression levels for each of the functional OR genes is shown. B6 expression (blue) is lower than 129 (yellow) and CAST (red).

Thus, the differences in expression observed between the different strains of mice analysed here are likely to be the product of genomic variation located in important regulatory elements, both in the promoters of the OR genes and in long-acting enhancer elements.

4.4 The genetic background determines OR expression levels independent of odour environment.

Genetic variation will, undoubtedly, have an effect on the expression of the OR genes. However, genetically divergent mouse strains also produce different odours in their urine[326, 327] and amniotic fluid[51]. Therefore each strain of mouse, when housed in homogeneous groups, is exposed to a unique pre- and post-natal olfactory environment. Odour exposure has been shown to alter the life-span of OSNs in an activity dependent manner[286, 288, 292]. Thus, the observed changes in OR gene expression between the different strains could be the result of their genetic makeup and/or the odour environment they are exposed to. I therefore designed an experiment to dissect the genetic from the environmental contribution to OR expression regulation.

B6 and 129 embryos were transferred to F1 mothers and allowed to develop in this equivalent *in utero* environment. After birth, B6 litters were cross-fostered to B6 mothers and 129 litters to 129 mothers. Further, B6 litters received a single 129 pup, and 129 litters received a single B6 pup. In this setting, each litter has a characteristic olfactory environment, but one animal (the *alien*) has a different genetic background (Figure 4.12). This arrangement was maintained after weaning into single-sex groups, such that each alien was caged with four animals of the alternative strain. At 10 weeks of age, the WOM was collected for six alien animals and six cage-mates and RNA-sequenced (Tables B.1 and B.2 in Appendix B).

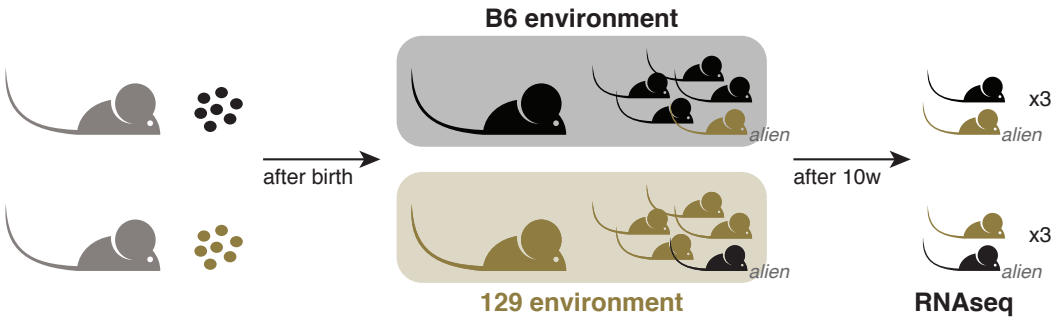

Figure 4.12 – Experimental design to dissect genetics from environment. B6 (black) or 129 (brown) embryos were transferred into F1 recipient mothers (grey). Once the pups were born, the B6 litters were cross-fostered to B6 dams and the 129 litters to 129 mothers. At the same time, a single B6 animal (the *alien*) was introduced to the 129 litter and *vice versa*. These groups were conserved after weaning, in 4:1 strain proportions. At 10 weeks of age, the WOM was collected from the alien animal along with a cage-mate. Three biological replicates were performed for each setup.

Evaluation of the OR gene expression repertoire revealed that animals clustered in two groups, clearly defined by the genetic background of the animals (Figure 4.13A). In other words, all the B6 animals displayed a similar OR gene profile, irrespective of their olfactory environment. Consistent with this, the correlation coefficients for any two B6 samples was on average 0.97, with no significant difference between the environments (p-value = 0.09, t-test); in contrast, the correlations for any B6 with a 129 sample had a mean of 0.89, which is significantly lower (p-value = 3.8e-12, t-test). 507 OR genes, among 5,475 genes were DE between these mice when grouped by strain, a similar number to the B6 vs 129 comparison when in their own olfactory environments (Figure 4.6). In striking contrast, across the whole transcriptome, only two genes showed differences in expression according to environment, both of which were OR genes (Figure 4.13B-C). For one of these ORs, the B6 animals in the 129 environment showed an expression pattern that resembles that of the 129 genome (Figure 4.13B); the other, however, had similar expression levels for both strains in their cognate environments,

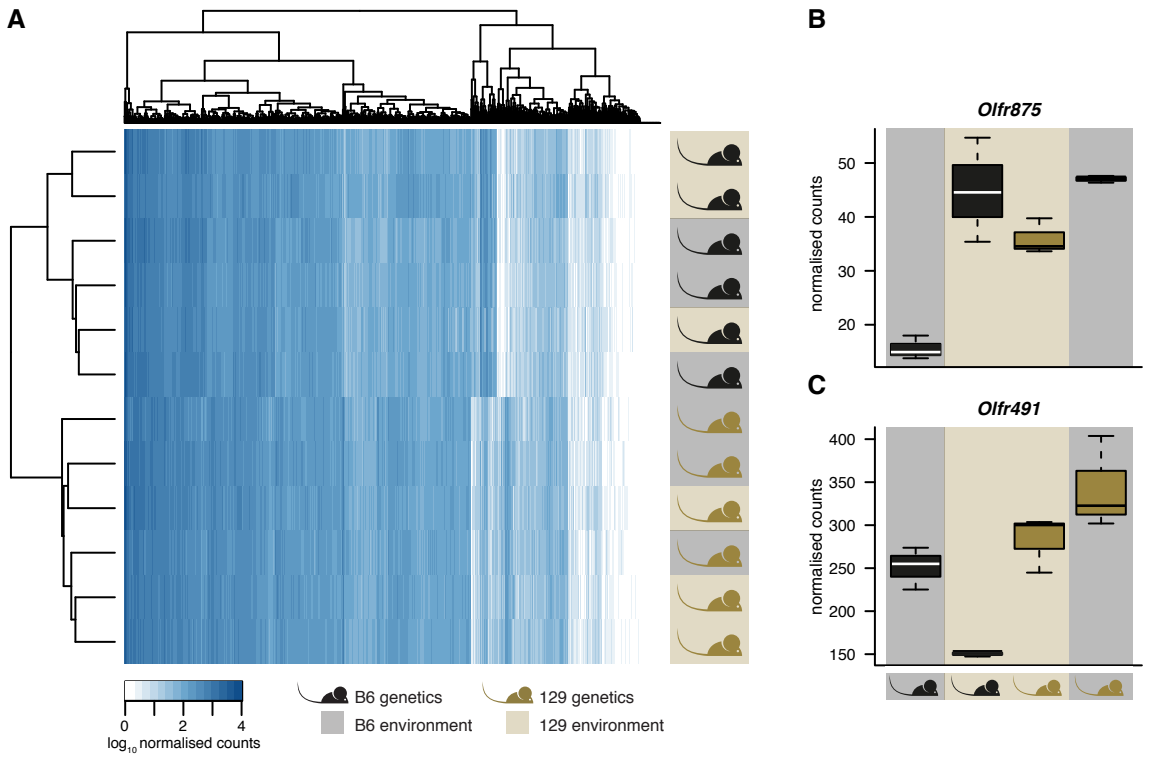

Figure 4.13 – OR expression is determined by the genetic background. A) Heatmap of the normalised OR expression for all the B6 (black) and 129 (brown) animals. To the right, the colour of the mouse indicates its genetic background and the colour surrounding it the type of environment it was exposed to. The samples cluster in two well defined groups, characterised by the strain of the animals. There is no clustering between the different environments. **B)** Two OR genes that showed altered expression levels upon differences on the olfactory environment. The colour of the box indicates the strain of the animal and the background shade the environment.

but when the B6 animals were switched to the 129 environment, the expression was downregulated (Figure 4.13C). No other genes reached statistical significance suggesting that environmental changes have very little influence on the regulation of gene expression in the main olfactory system. These data demonstrate that the WOM transcriptome is mostly influenced by direct genetic effects and the indirect effect of the olfactory environment is minimal and perhaps restricted to only a couple of OR genes. What's more, these results imply that the overall abundance of each OR type is independent of its activity or responsiveness to odorant cues.

To further test this, I sequenced the WOM of newborn B6 pups[1]. These animals were mostly restricted to their *in utero* environment and had only brief interaction with outside odorants. Therefore, the observed expression of the OR genes had minimal influence from olfactory stimulation. The expression of OR genes in these pups could

[1]Tissue collected by Darren Logan.

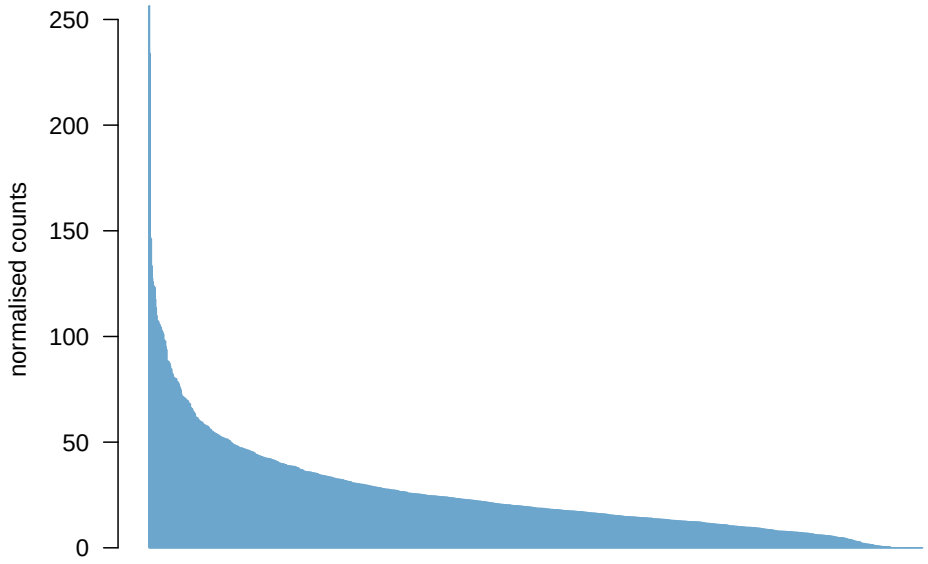

Figure 4.14 – OR expression in B6 pups. OR expression levels in B6 E19.5 newborn animals. Despite the abundance of OR expression being lower than in adults, there is a clear distribution of unequal proportions of the different OR genes, with some much more abundant than others.

be readily detected in the RNAseq data. From the whole OR repertoire, 1,198 (95.9%) genes showed evidence of expression. The olfactory system is still developing at birth and there is rapid growth and a steep increase in the number of OSNs during the first postnatal weeks[26]. Consistently, the expression levels of the receptor genes were low compared to adult animals. Nonetheless, a similar distribution of unequal proportions of different OR genes was readily observed (Figure 4.14). The median OR expression was 18.76 normalised counts, and only 24 receptor genes had abundances higher than 100 normalised counts. Thus, the differential proportions of OSNs expressing particular OR genes are present before any significant olfactory stimulation has occurred, suggesting that it is not dependent on the activity of the OSNs.

Two other observations support that the expression distribution observed is not influenced by the activity of the receptor genes. First, analysis of the expression of receptors that are pseudogenised and do not produce functional proteins revealed a similar distribution, of some genes expressed more abundantly than others (Figure 4.15). And second, the collection of OR genes that have identical coding sequences between different strains –and therefore identical proteins– very often occupy varying positions in the distribution. 36.3% of the OR genes that are identical between B6 and 129 are significantly DE; similarly 44.8% of the identical receptors between B6 and CAST are expressed at different levels.

Thus, the activity of the receptor protein itself has no influence on the final expression

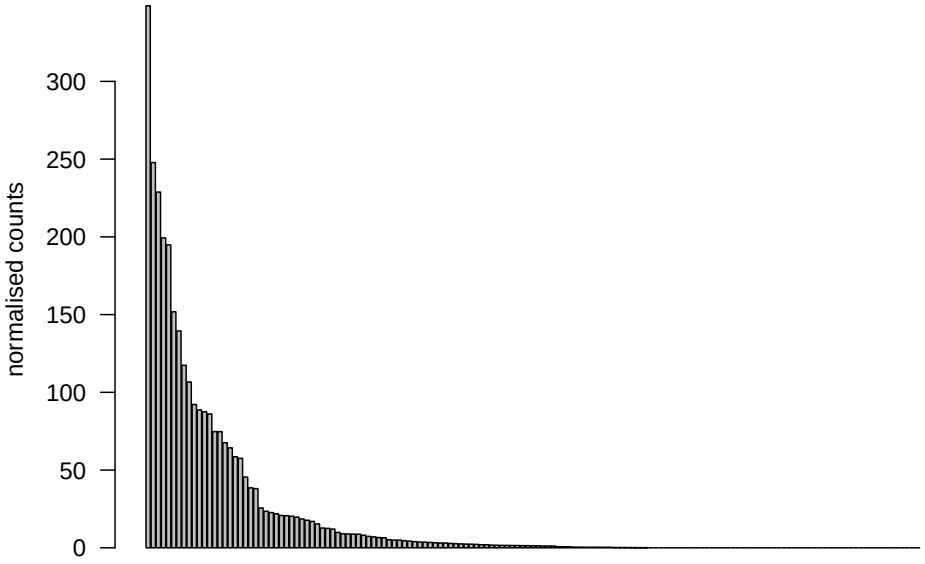

Figure 4.15 – Pseudogene OR expression in B6. OR expression levels for all those receptors annotated as pseudogenes in the B6 genome. Some genes are expressed at much higher levels than others, suggesting that the abundance of each receptor is independent of the activity of the protein.

level of the gene. All data indicates that the observed differences in OSN composition in the MOEs of different mouse strains are driven by changes in their genetic architecture; the state of the regulatory sequences present in each genome dictate the final proportion of OSNs that express each receptor type.

4.5 OR expression is controlled in *cis*.

Several enhancer elements have been identified that regulate the probability with which OR genes from nearby clusters are chosen[183, 186, 187]. For these, it has been demonstrated that their regulatory activity acts in *cis* and do not influence the expression of the homologous alleles on the other chromosome[183, 184, 186]. However, 3C experiments have indicated that there are interchromosomal interactions between different enhancer elements[187]. Further, I have now shown that the expression level of each receptor gene depends on the genetic context. In order to determine if the observed differences in expression are the product of *cis*-acting elements, I analysed available WOM data from B6×CAST F1 hybrids[2]. When compared to the parental strains, these hybrid animals provide information about the regulatory elements affecting gene expression. For all those genes that are differentially expressed between the parental strains, the corres-

[2]Raw sequencing data was kindly provided by Sophia Liang.

ponding expression of each allele can be determined in the F1 by interrogating variable positions. If the ratio observed between the parental expression levels is the same as that of the two alleles, it can be inferred that expression is regulated by elements acting in *cis*; on the other hand, if the expression of the two alleles in the F1 is no longer different, then regulation is occurring in *trans*[328].

For the OR repertoire, 1,018 (81.5%) of the genes have at least one SNP reported for CAST that is covered by sequencing fragments. At each of these positions, I calculated how many sequencing fragments pertained to each allele. Then, I used the ratio of B6 to CAST counts to deconvolve the total gene expression into allele-specific expression for each OR gene (see Appendix A for detailed methods). Figure 4.16A shows the fold-change in OR gene expression in the parental strains versus the corresponding fold-change of the alleles in the F1; most genes lie along the 1:1 diagonal, indicating that the expression levels observed in the animals with pure genetic background (B6 or CAST) are conserved at the allele-level in the F1 and, therefore, must be controlled in *cis*. In other words, the expression of the two alleles within the F1 animals is the summation of the patterns present in the parents (Figure 4.16B).

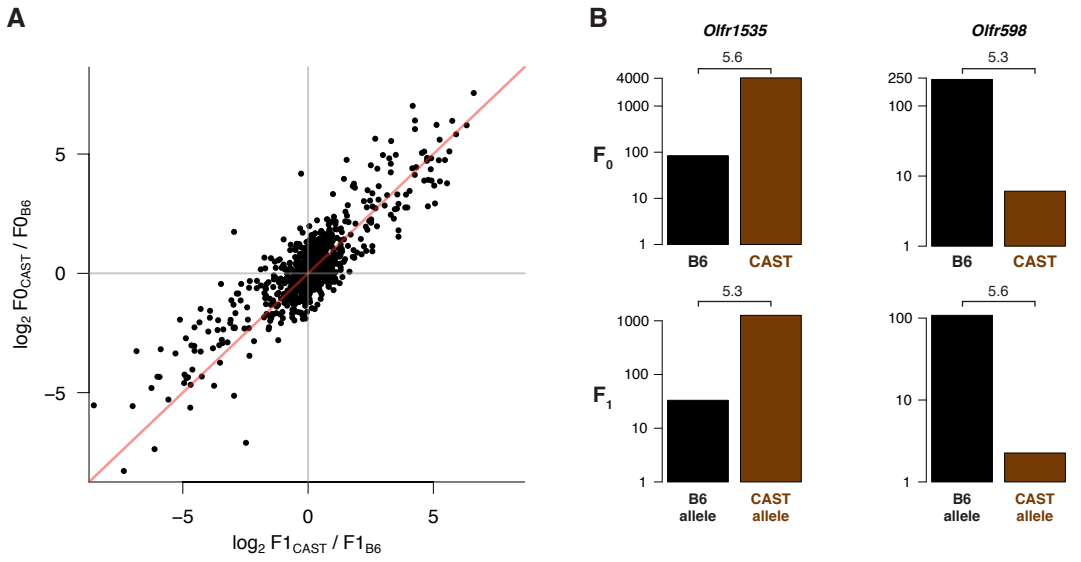

Figure 4.16 – OR expression is regulated in *cis*. A) Scatter plot of the fold-change in OR expression between the CAST and B6 parental strains (y-axis) versus the allelic expression for each strain in the F1 (x-axis). Most genes lie in the 1:1 diagonal (red), which indicates that the proportions observed in the parents are conserved in the F1. This occurs when expression is regulated by *cis*-acting elements. **B)** Examples of two OR genes that are differentially expressed in the parental strains (top row). The expression in the F1 hybrids was decomposed into the contribution from each allele, and this is plotted on the bottom row. The $\log_2$ fold-change is indicated in each case and these are equivalent for both the F_0 and F_1 data.

All together, these data provide a comprehensive landscape of the OR repertoire transcriptome from genomes with varying levels of divergence. A great proportion of

the receptor genes are susceptible to differences in their expression levels. These correlate with disparity in the number of OSNs that express each OR gene. Further, the changes are directly determined by the genetic architecture of the animal but not by their olfactory environment or the activity of the receptor protein. Based on the expression pattern of OR genes in an F1 hybrid, it can be inferred that the vast majority of regulation of OR gene choice is conducted by elements acting in *cis*. Genetic variation within these –poorly defined– regulatory elements correlates with differential OR gene expression. Together, these data are consistent with a model where genetic variation in regulatory elements alters the probability with which each receptor is chosen, thus resulting in the creation of a highly diverse mosaic composition of the MOE, with varying proportions of each OSN type.