Published in final edited form as:

Trends Cell Biol. 2013 September ; 23(9): 449–456. doi:10.1016/j.tcb.2013.04.005.

Stochastic gene expression in mammals: lessons from Olfaction

Angeliki Magklara¹ and Stavros Lomvardas²

¹Department of Anatomy, University of California San Francisco CA 94920, USA

²Division of Biomedical Research, Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, Ioannina, Greece

Summary

One of the remarkable characteristics of higher organisms is the enormous assortment of cell types that emerge from a common genome. The immune system, with the daunting duty of detecting an astounding number of pathogens, and the nervous system with the equally bewildering task of perceiving and interpreting the external world, are the quintessence of cellular diversity. As we began to appreciate decades ago, achieving distinct expression programs among similar cell types cannot be accomplished solely by deterministic regulatory systems, but by the involvement of some type of stochasticity. In the last few years our understanding of these non-deterministic mechanisms is advancing, and this review will provide a brief summary of the current view of stochastic gene expression with focus on olfactory receptor gene choice, the epigenetic underpinnings of which recently began to emerge.

Keywords

stochasticity; olfactory receptors; clustered protocadherins; antigen receptors; epigenetic mechanisms; nuclear architecture

Stochastic decisions in gene expression

Stochasticity in gene expression refers to the random mechanisms that govern transcription and translation resulting in variable levels of mRNA and proteins across cells of the same population. Stochastic gene expression has been studied mainly in prokaryotic organisms and lower metazoan, where it provides the means for genetically identical populations to obtain phenotypic diversity and develop subpopulations with adaptive advantages that can be used for their survival in varying environments [1]. One would expect that stochastic gene expression would not be tolerated in higher eukaryotes, where complex regulatory circuits control reproducible differentiation patterns that have prevailed evolutionarily [2]. Nonetheless, a new wave of studies revealed that stochastic choices are often found at the basis of central developmental programs dictating important cell-fate decisions or inducible transcriptional choices [3, 4]. Several excellent reviews have addressed the stochastic mechanisms of cell-fate specification in various organisms and the principles that connect stochastic gene expression with the biological functions it enables [5–8].

© 2013 Elsevier Ltd. All rights reserved.

Correspondence to: Stavros Lomvardas.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

In this review, we use the regulation of the olfactory receptor genes as a paradigm of stochastic, yet irreversible, gene expression decision. The selection of a single olfactory receptor (OR) allele in each olfactory sensory neuron (OSN) in the mouse olfactory epithelium (OE) is probably the longest studied example of stochastic choice in the mammalian nervous system. We provide an overview of models proposed to explain OR stochastic expression and we describe newly identified aspects of the OR genes' complex regulation that highlight the critical role of epigenetic mechanisms, locus repositioning and nuclear architecture as contributing factors.

The olfactory system is often paralleled to the immune system, as they share several characteristics including monoallelic (see **Glossary**) receptor expression. As a matter of fact, allelic exclusion (see **Glossary**) of the antigen receptor genes is one of the first and most thoroughly studied cases of stochastic choice. Therefore, we begin this review by summarizing key findings on the monoallelic expression of antigen receptors in lymphocytes; we proceed to describe in detail mechanisms that mediate stochastic choices in OR expression and discuss analogies between the two systems. Finally, we present recent developments in the expression of clustered protocadherins that confirm the importance of epigenetic silencing and long range DNA interactions in stochastic processes.

Stochasticity in the immune system: allelic exclusion of antigen receptor genes

As early as in the mid-60's there was evidence of the monoallelic nature of the antigen receptors' expression [9]. Later studies established that allelic exclusion occurs during the V(D)J recombination process (BOX I) and only one allele is successfully rearranged and expressed. Even though the nature of allelic exclusion, probabilistic versus deterministic, is still a matter of debate, it is notable that both schools of thought evoke stochastic choices to account for this phenomenon [10]. The probabilistic model describes the monoallelic rearrangement as the result of random choice between two equivalent alleles and as the consequence of the low probability of simultaneous efficient recombination. Experimental support for this model was provided by the observation that both alleles of the TCR β locus were distributed frequently and stochastically within the nuclear lamina and pericentromeric heterochromatin foci [11]. Notably, these alleles were less likely to undergo V β -to-D β J β recombination in Double Negative thymocytes [11]. Therefore, it was proposed that stochastic interactions with repressive nuclear compartments could reduce the likelihood of simultaneous VDJ recombination [12]. The deterministic (instructive) model favors an initial stochastic marking of one allele at an early developmental stage; this marking is clonally maintained and ultimately dictates (instructs) the successful rearrangement and activation of the associated locus. A hallmark of monoallelic expression, asynchronous replication (see **Glossary**), was determined to set the mark for the subsequent allelic exclusion of antigen receptor genes [13]. A recent study has shed more light into this mechanism [14]. Using the Ig κ locus as a model system, the authors showed that commitment to a chosen allele occurs in early lymphoid lineage cells, it is accompanied by changes in asynchronous replication, it is clonally maintained and it predetermines monoallelic rearrangement in B cells [14]. This model has gained wider acceptance and describes allelic exclusion as a phenomenon that evolves in two phases, initiation and maintenance. The initiation phase, apart from the stochastic asynchronous replication, involves additional layers of regulation that ensure only one productive recombination will occur at each antigen receptor locus. Such mechanisms include the preferential association of the late replicating antigen allele with pericentromeric heterochromatin and the monoallelic contraction by chromosome looping that leads to the juxtaposition of V and D-J segments [15], as well as changes in the DNA methylation and the histone modification status of the

replicating allele that allow it to be accessible to recombination enzymes [16–18] (Figure 1a). In the maintenance phase, once an in-frame rearrangement takes place, a feedback mechanism is elicited that inhibits further recombination and is associated with epigenetic and locus conformation changes. For example, the non-functional IgH allele gets recruited to pericentromeric heterochromatin and adopts a closed chromatin state [19] (Figure 1b). These changes could inhibit recombination by preventing the binding of recombinases to the unrearranged loci. In addition, unrearranged IgH alleles, although contracted in pro-B cells, become decontracted in pre B-cells, separating the V segments from the DJ segments and decreasing the likelihood of a new allelic recombination [19].

BOX I

There are seven distinct and structurally unique antigen receptor loci, the heavy chain locus (IgH), the light chain loci (Ig κ and Ig λ) and the T cell receptor -alpha (TCR α), -beta (TCR β), -gamma (TCR γ) and -delta (TCR δ) loci. They are composed of multiple Variable (V), Diversity (D) and Joining (J) gene segments as well as of the constant (C) exons. Example of the organization of an antigen receptor locus is shown in Figure I. The mouse IgH locus covers ~3MB and consists of ~150 Variable (V_H), 9–12 Diversity (D_H) and 4 Joining (J_H) gene segments. C μ and C α are the constant exons (locus is not drawn to scale). The V(D)J recombination process involves the random rearrangement of the homonymous gene segments in order to generate the variable part of the immunoglobulin and T-cell receptor genes. This process is cell lineage restricted and developmentally-dependent. In the B-lineage, the IgH locus is the first to be rearranged with a D segment joining a J segment (pre pro-B cells) followed by V to DJ recombination (pro-B cells) (Figure I). Light chain gene (Ig κ or Ig λ) rearrangement occurs in the next developmental stage of pre-B cells. The Ig κ and Ig λ loci do not contain D segments and recombination occurs only between the V and J segments (Figure I). Similarly, in the T-lineage, the T cell receptor (TCR) gene segments undergo the same sequence of ordered recombination events. The TCR β locus is firstly rearranged through D-to-J recombination, which is followed by V-to-DJ rearrangement (Double Negative thymocytes). In the next stage of T-cell maturation (Double Positive thymocytes) the rearrangement of the TCR α locus occurs through joining of V-to-J segments.

Figure I.
Genomic organization and recombination in the immunoglobulin family

Allelic exclusion ensures the monospecificity of lymphocytes, a property of the immune system critical for the detection and discrimination of different antigens and the subsequent

activation of appropriate responses. In the case that the B cells, for example, could express both alleles of the heavy and light chain loci, then several different immunoglobulins would be present on the cell surface. Activation of a B cell by an antigen would lead to upregulation and secretion of all of them, diluting the effective dose of the antigen-targeted antibody and/or triggering unwanted reactions (e.g. allergies, autoimmunity) [19].

Stochasticity in the nervous system: the OR choice paradigm

In analogy with the immune system, where the “one antigen receptor-one lymphocyte” rule guarantees proper function, the correct wiring of the olfactory system is “built” on the “one olfactory receptor-one neuron” rule.

The OR genes comprise the largest mammalian gene family and is particularly multitudinous in mouse (*Mus musculus*), where it consists of ~1,400 members scattered in more than 40 clusters throughout the genome [20]. The ORs are expressed almost exclusively in the main olfactory epithelium (MOE) that lines the nasal cavity, where they are responsible for odor detection. A remarkable characteristic of this family is that only one OR allele out of hundreds is expressed in every OSN [21]. Each expressed OR is intimately interconnected with the neuron’s identity, as it determines the particular odorant that will stimulate it, but it also instructs the targeting of the neuron’s axons to a specific glomerulus (see **Glossary**). Expression of more than one OR allele would disrupt the proper stimulation and wiring of the olfactory system and may lead to misinterpretation of chemical signals normally translated in the brain to the sense of smell.

As soon as the monogenic (see **Glossary**) and monoallelic nature of OR expression was revealed, the deciphering of the selection process that leads to activation of one allele and silencing of the rest became a subject of enormous interest in the field. Initially, two main models were proposed in accordance with the two “extreme” principles that underlie gene regulation: determinism and stochasticity. Determinism mandated the existence of ~1,400 distinct combinations of transcription factors, each combination acting on unique *cis*-elements for each OR gene (reviewed in [22]). The presence of at least 4 stereotypic zones of OR expression within the MOE, together with the fact that the frequency of expression of each OR is reproducible from mouse to mouse, provided support for this model. However, early experimental evidence, showing that a transgene and an endogenous OR allele that share identical regulatory sequences are never co-expressed, argued against it (reviewed in [22]). Moreover, the discovery of gene switching [23] (see **Glossary**) showed that a neuron had the regulatory potential to express multiple ORs. Consequently, a stochastic model allowing random expression of a subset of the OR repertoire within a zone has become the prevailing view for the monogenic and monoallelic OR gene expression. Although many models have been proposed, as described below, this regulatory mechanism remains, by and large, elusive.

Singularity of OR choice

As the monoallelic nature of OR expression was reminiscent of allelic exclusion in lymphocytes, a first attractive model to explain single OR choice suggested that irreversible DNA changes (similar to the recombination of antigen receptor genes) were involved. It had been, previously, shown that monoclonic mice generated from mature B- or T- lymphocytes had only the respective receptors rearranged in all tissues [24]. A similar task was undertaken by two groups, who cloned mice from post-mitotic olfactory neurons expressing specific ORs [25, 26]. They reasoned that if DNA rearrangements were involved in OR choice, the mice would display restricted OR expression. However, the cloned mice expressed the whole repertoire of ORs and the DNA structure around the selected OR loci

was intact, refuting the model of irreversible DNA rearrangement as a method for OR gene activation [25, 26].

A second model that attempted to explain the singularity of OR choice stemmed from elegant experiments in the visual system, which showed that a single enhancer could select one of the two opsin genes residing on the X chromosome (reviewed in [5]). In a similar fashion, it was proposed that each OR cluster (see **Glossary**) had its own enhancer that physically interacted with one of the proximal OR alleles. The discovery of the H enhancer [27] introduced the notion of distant OR enhancers, or LCRs (see **Glossary**), as the mediators of singularity in OR choice. A ~2 Kb sequence, which is highly conserved between mouse and human, was isolated 75 kb upstream of a small cluster of OR genes. It was shown that the H enhancer was required for the transgenic expression of these OR genes suggesting that stochastic interaction of this element with a single OR promoter *in-cis* could lead to monoallelic activation of the specific gene. A second OR enhancer that was recently identified, the P element, also affects expression of a few neighboring OR alleles [28], further supporting the idea that local enhancer sequences may be necessary for the activation of an OR. The deletion of H and P, however, affect the expression of only a small number of OR genes, leading to the estimation that about 200 similar local enhancer elements might exist in the mouse genome and be sufficient for the expression of the whole OR repertoire.

A third model, which is not mutually exclusive to the LCR hypothesis, suggests that nuclear architecture and locus repositioning play an important role in the singularity of OR choice. An explosion of genome-wide data, ignited by the seminal development of the chromosome conformation capture (3C) assay, demonstrates that the genome is not randomly distributed in the nucleus, but is organized in chromatin territories (reviewed in [29]). Moreover, gene transcription occurs in well-defined nuclear factories, where co-regulated genes cohabitate, in a mechanism thought to afford synergistic activation and more efficient coordination of transcriptional responses (reviewed in [30]). In accordance with these findings, OR genes were shown to be highly organized in the OSN nucleus [31]. The majority of silent alleles converge to a small number of OR-specific heterochromatic foci, whereas active OR alleles reside in proximal but distinct euchromatic territories. Genetic manipulations disrupting OR aggregation results in co-expression of a large number of OR genes in each OSN, and significant down-regulation of OR transcription [31]. This suggests that escape from repressive OR foci is not the only requirement for robust OR transcription, but also relocation to a specialized transcription factory might be a necessary second step for the completion of this process [31]. In support of this, the active OR allele in each OSN frequently interacts, in *cis* or *trans*, with the H enhancer [32]. The fact that the H enhancer is required only for the expression of three proximal ORs [27], despite its physical association with many more OR genes [32], may reflect the existence of an OR-specific transcription factory defined by the presence of H and potentially other OR enhancers that cooperate for OR activation, as it was recently shown for the Hox [33] and Protocadherin gene clusters [34]. Indeed, disruption of OSN nuclear architecture abolished interactions between H and ORs in *trans* [31], supporting the notion that massive nuclear reorganization may have ablated an OR-specific transcription factory resulting in down-regulation of OR transcription.

Stabilization of OR choice

Although little is known about the mechanisms that ensure only one OR allele is selected for transcriptional activation in each OSN, much more is understood about the process that preserves this singular expression. New ground was broken by the realization that transgenic ORs elicit a negative feedback signal that prevents the co-expression of endogenous OR alleles [27]. This feedback depends upon the expression of intact, full-length OR protein,

since transgenes that lack the OR coding sequence (CDS) or carry a premature stop codon cannot prevent the expression of endogenous ORs [27]. The OR CDS appears to be important also for the ability to receive that signal, since transgenes containing an OR CDS are expressed in a higher percentage of OSNs, if their transcription is initiated before, rather than after, the onset of endogenous OR expression. Although this negative feedback requires expression of the full length OR protein, it is independent of the ability of the OR to activate its signaling pathway, since mutant transgenic ORs that cannot interact with G-proteins, retain their singular expression pattern [35]. In addition to this negative feedback, lineage tracing experiments from endogenous OR loci revealed the existence of a positive feedback loop that stabilizes the expression of the chosen allele [23]. This signal also relies on the production of full length OR protein.

The observation that OR production elicits a signal with two distinct effects, one to stabilize the robust expression of the chosen OR and the other to prevent the transcriptional activation of additional alleles, suggests that the target of this signal may be an activity that is required for the initiation of OR transcription but dispensable for its stabilization. Recent characterization of the epigenetic state of silent and active ORs in the MOE provides a blueprint of the pathway leading to OR activation and candidate targets for a feedback signal with dual activity [36]. Specifically, genome-wide chromatin analysis of the MOE revealed a novel epigenetic signature for OR genes consisting of H3K9me3 and H4K20me3 [36], a surprising finding, since these histone modifications are primarily associated with constitutive heterochromatin in pericentromeric and subtelomeric repeats (see **Glossary**). These repressive epigenetic marks are not restricted to the promoter sequence but extend throughout the OR genes and their intergenic regions, generating large heterochromatic blocks that cover specifically all of the OR genomic clusters [36]. This spreading of OR heterochromatin likely serves regulatory purposes by masking transcription factor binding sites enriched on OR promoters, which also extend for hundreds of base pairs upstream and downstream of the OR transcription start sites [37]. The observation that a reporter transgene, driven by a heterologous pan-OSN promoter, is expressed in a sporadic, monogenic and likely monoallelic fashion, when inserted within an OR-flanking heterochromatic block [36], suggests that this unique epigenetic signature may play an instructive role in OR regulation and provides a striking example of epigenetic regulation overriding genetic information encoded in a promoter.

The presence of H3K9me3 and H4K20me3 on OR loci before the last cell division, and prior to terminal neuronal differentiation, indicates that the OR genes are silenced long before the onset of OR expression, which occurs during the maturation of post-mitotic OSNs. This finding, together with the demonstration that the expressed OR allele lacks these repressive modifications and is instead marked by H3K4me3 [36], suggests that OR activation coincides with, and likely depends upon, an epigenetic switch from a heterochromatic to a euchromatic epigenetic signature. This epigenetic switch appears specific, albeit stochastic, since neighboring OR alleles, or the identical OR allele that is inherited from the other parent, retain the repressive histone marks [36].

A “two-step” model for stochastic OR choice

A model that incorporates the data described predicts that the singularity of OR expression is achieved in two phases, similarly to the allelic exclusion of antigen receptor genes: a) an activation process that results in the expression of a single OR allele and b) a maintenance phase that prevents the activation of additional OR alleles and stabilizes the expression of the initially chosen OR (Figure 2).

Before the activation phase, all OR genes are already repressed and are found restricted in heterochromatic nuclear foci. It is possible that the epigenetic state of the OR genes functions as a code recognized by a limited enzymatic activity that removes the repressive methylation marks from a stochastically chosen allele (Lyons et al., unpublished); subsequently, this is moved outside the heterochromatic foci in a transcriptionally permissive environment (Figure 2a). Interaction with a specific enhancer (or enhancer complex) may be necessary for transcriptional activation and may contribute to the singularity of OR choice. This is also achieved by the epigenetic silencing of the ORs before the onset of OR transcription making their transcriptional activation a slower and more complex process.

The maintenance phase consists of a feedback signal elicited by the functional OR protein that does not silence the non-chosen ORs *per se*, but it targets the histone demethylases that had removed the repressive histone modifications from the activated OR allele (Lyons et al unpublished) (Figure 2b). This feedback mechanism does not affect the transcription of the already de-repressed allele (Lyons et al unpublished), located outside the OR foci, but it prevents the transcriptional activation of the remaining ORs residing in the heterochromatic foci (Figure 2b).

Stochastic promoter activation in clustered protocadherins

Although OR choice presents an extreme regulatory challenge, due to the large number of family members and their clustered distribution in many chromosomes, it is not the only example of stochastic gene choice in the nervous system. Important principles for stochastic and monoallelic gene expression emerge also from the study of clustered protocadherin gene expression. Clustered protocadherins (Pcdh α , β and γ) play an important role in interneuronal recognition and dendritic self-avoidance [38]. The genes are organized in tightly linked genomic clusters that span ~1MB. The Pcdh α and γ genes contain multiple 'variable' exons that encode for the protein's extracellular domain, as well as, a set of 'constant' exons that encode for the common intracellular region [39] (Figure 3a). Each variable exon has its own promoter and when activated it is spliced to all three constant exons (Figure 3b). Through *cis*-alternative splicing a large number of transcripts can be generated by each gene cluster (Figure 3c). In individual neurons, the Pcdh- α , - γ A and - γ B isoforms are expressed in an unusual monoallelic and combinatorial pattern. Even though both the maternal and the paternal allele are expressed, the stochastic activation of a different promoter from each cluster leads to the expression of different sets of protocadherins in each cell. In contrast, all the C-type isoforms are biallelically expressed [40].

The mechanisms that mediate the stochastic choice of one promoter are currently unknown. Two tissue specific enhancers, HS7 and HS5-1 (HS: hypersensitive sequence) that play a role in the activation of the variable exons have been identified in the Pcdh α cluster (Figure 3) [41]. Deletion of HS5-1 in neuronal tissues results in down-regulation, but not complete elimination, of expression of some variable exons of the alpha cluster [34], while, in non-neuronal tissues, it results in up-regulation. Deletion of HS7, on the other hand, causes a moderate but uniform down-regulation of Pcdh α expression in a subset of brain regions [34]. Therefore, it is possible that stochastic promoter activation in this system is mediated via its interaction with a distant regulatory element, such as HS7 and HS5-1. Recent work has also revealed an epigenetic component in this process. One study showed that CTCF and cohesin preferentially bind to transcriptionally active promoters, as well as to the enhancers of the Pcdh α cluster [42]. Experiments utilizing the 3C technique confirmed that active promoters and enhancers are brought together by CTCF/cohesin complexes through long-range DNA looping [43] and that DNA methylation of the promoters prevents CTCF

binding. It is known that the silenced protocadherin alpha promoters are heavily methylated and treatment with demethylation agents causes expression of all isoforms [44]. In olfactory sensory neurons, the protocadherins are marked by H3K9me3 and H4K20me3 in the silent variable exons, but not in the constant exons [36]. One potential model to explain this is: i. Initially all the variable promoters are unmethylated and CTCF/cohesin bound (Figure 3d); ii. Stochastic DNA-looping brings the promoter that will be activated close to the enhancers and stabilizes the complex in a transcriptional hub [43] (Figure 3e); iii. The other variable promoters eventually lose CTCF/cohesin binding and gain DNA hypermethylation, “locking” them in a permanent silent state, which is further reinforced by heterochromatinization (Figure 3e).

Concluding remarks

The aforementioned examples of stochastic gene expression represent variations of the same strategy to create transcriptional diversity among large mammalian gene families. This strategy features the use of epigenetic silencing as a means to limit the number of available genes, and the utilization of long-range DNA interactions as a way to restrict the number of alleles that could be activated in each cell. Depending on the role of each gene family and the biological needs of the cells that express them there are mechanistic variations in the regulation and stabilization of choice. In the case of immunoglobulin and T-cell receptor choice, which are used for the detection of an astounding number of constantly evolving pathogens, epigenetic silencing and long-range genomic interactions are used to achieve singularity of allele choice, followed by DNA rearrangements that assure maximum diversity and may also contribute the preservation of a properly rearranged locus for the life of the lymphocyte. In the case of Pcdhs and ORs, there is no need for such an extreme measure for diversity and irreversibility. A repertoire of 1000 or so OR genes is probably the optimal number of receptor diversity for robust neuronal computations in the olfactory bulb and higher processing centers. Similarly, the existing number of variable Pcdh genes in the three genomic clusters can generate over 3 million different combinations or “barcodes”, which is sufficient for the distinction of self vs. non-self in a local neurite milieu. Moreover, stabilizing OR and Pcdh choice over long periods of times may not be as a pressing requirement as in memory B and T lymphocytes, which ought to express the same receptor for the life of the organism. OSNs have an average lifespan of 3 months, and a change in Pcdh repertoire over time may not prohibit self-repulsion. For this reason, it is possible that Pcdh expression does not elicit any type of specific feedback to stabilize the expression of the chosen alleles, whereas in the case of OR choice the elicited feedback preserves the epigenetic state of the active and silent OR alleles (Lyons et al., unpublished). Notably, it was recently demonstrated that in addition to the enigmatic OR-elicited feedback signal, there is a second step of quality control that reduces the longevity of OSNs that express infrequently used ORs [45], a process that is reminiscent of the elimination of self-reacting or non-functional lymphocytes during positive and negative selection [46, 47].

Beyond these common characteristics and differences, however, there is a bigger conceptual challenge for all three gene families; how to achieve singularity of initial choice. Even in the case of Pcdh choice, where all the genes reside on one chromosome, the answer is not simple, since there are more than one *cis* enhancers implicated in this process. The problem becomes exponentially more complicated in the case of OR choice, since OR genes can be essentially found on every chromosome. Thus, deciphering the mechanisms of singular gene choice was, and shall remain, the holy grail of stochastic gene expression.

Glossary

Monoallelic expression	Usually both alleles of a gene are actively transcribed (bi-allelic expression). In some cases, only one allele of a gene is expressed (monoallelism), for example several X-linked genes in females due to X-chromosome inactivation. Recent studies have revealed that quite a large number of autosomal genes display random monoallelic pattern of expression [8]
Allelic exclusion	It refers to the expression of a single allele from a specific gene locus and silencing of the other allele. In lymphocytes this phenomenon leads to the expression of one type of antigen receptor per cell
Asynchronous replication	It is a phenomenon frequently observed with autosomal monoallelically expressed genes, imprinted genes and X-linked genes, where one allele is replicated earlier than the other. In most genes both alleles are replicated at the same time at a specific point in S phase
Glomerulus	A specific structure in the olfactory bulb where all the olfactory sensory neurons that express the same olfactory receptor project their axons to. The glomeruli consist of the synapses of the OSNs' axons with the mitral cells and they form an olfactory topographic map that makes possible the interpretation of transmitted chemical signals to the brain as different odorants
Monogenic expression	It refers to the expression of a single gene or a pair of allelic genes
Gene switching	A process where neurons that selected a non-functional (and sometimes functional) OR re-choose a different OR allele
OR cluster	The vast majority of OR genes in mouse are found in groups (clusters) of two to several dozens of genes scattered throughout the genome
Locus control regions	are “defined by their ability to enhance the expression of linked genes to physiological levels in a tissue-specific and copy number-dependent manner at ectopic chromatin sites. The components of an LCR commonly colocalize to sites of DNase I hypersensitivity (HS) in the chromatin of expressing cells. The core determinants at individual HSs are composed of arrays of multiple ubiquitous and lineage-specific transcription factor-binding sites” [48]
Pericentromeric and subtelomeric repeats	Highly repetitive DNA regions found adjacently to the centromere and the telomeres of chromosomes. They are associated with DNA hypermethylation and the histone modifications H3K9me3 and H4K20me3

References

1. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*. 2008; 135:216–226. [PubMed: 18957198]
2. Johnston RJ Jr, Desplan C. Stochastic neuronal cell fate choices. *Curr Opin Neurobiol*. 2008; 18:20–27. [PubMed: 18511260]

3. Apostolou E, Thanos D. Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell*. 2008; 134:85–96. [PubMed: 18614013]
4. Zhao M, et al. Stochastic expression of the interferon-beta gene. *PLoS biology*. 2012; 10:e1001249. [PubMed: 22291574]
5. Johnston RJ Jr, Desplan C. Stochastic mechanisms of cell fate specification that yield random or robust outcomes. *Annu Rev Cell Dev Biol*. 2010; 26:689–719. [PubMed: 20590453]
6. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*. 2010; 467:167–173. [PubMed: 20829787]
7. Balazsi G, et al. Cellular decision making and biological noise: from microbes to mammals. *Cell*. 2011; 144:910–925. [PubMed: 21414483]
8. Chess A. Mechanisms and consequences of widespread random monoallelic expression. *Nature reviews*. 2012; 13:421–428.
9. Pernis B, et al. Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J Exp Med*. 1965; 122:853–876. [PubMed: 4159057]
10. Cedar H, Bergman Y. Choreography of Ig allelic exclusion. *Curr Opin Immunol*. 2008; 20:308–317. [PubMed: 18400481]
11. Schlimgen RJ, et al. Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nat Immunol*. 2008; 9:802–809. [PubMed: 18536719]
12. Krangel MS. Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol*. 2009; 21:133–139. [PubMed: 19362456]
13. Mostoslavsky R, et al. Asynchronous replication and allelic exclusion in the immune system. *Nature*. 2001; 414:221–225. [PubMed: 11700561]
14. Farago M, et al. Clonal allelic predetermination of immunoglobulin-kappa rearrangement. *Nature*. 2012; 490:561–565. [PubMed: 23023124]
15. Roldan E, et al. Locus ‘decontraction’ and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol*. 2005; 6:31–41. [PubMed: 15580273]
16. Sikes ML, Oltz EM. Genetic and epigenetic regulation of Tcrb gene assembly. *Curr Top Microbiol Immunol*. 2012; 356:91–116. [PubMed: 21688096]
17. Subrahmanyam R, Sen R. Epigenetic features that regulate IgH locus recombination and expression. *Curr Top Microbiol Immunol*. 2012; 356:39–63. [PubMed: 21779986]
18. Cedar H, Bergman Y. Epigenetics of haematopoietic cell development. *Nat Rev Immunol*. 2011; 11:478–488. [PubMed: 21660052]
19. Vettermann C, Schlissel MS. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol Rev*. 2010; 237:22–42. [PubMed: 20727027]
20. Zhang X, Firestein S. The olfactory receptor gene superfamily of the mouse. *Nat Neurosci*. 2002; 5:124–133. [PubMed: 11802173]
21. Chess A, et al. Allelic inactivation regulates olfactory receptor gene expression. *Cell*. 1994; 78:823–834. [PubMed: 8087849]
22. Shykind BM. Regulation of odorant receptors: one allele at a time. *Human molecular genetics*. 2005; 14(Spec No 1):R33–39. [PubMed: 15809271]
23. Shykind BM, et al. Gene switching and the stability of odorant receptor gene choice. *Cell*. 2004; 117:801–815. [PubMed: 15186780]
24. Hochedlinger K, Jaenisch R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature*. 2002; 415:1035–1038. [PubMed: 11875572]
25. Eggen K, et al. Mice cloned from olfactory sensory neurons. *Nature*. 2004; 428:44–49. [PubMed: 14990966]
26. Li J, et al. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature*. 2004; 428:393–399. [PubMed: 15042081]
27. Serizawa S, et al. Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science*. 2003; 302:2088–2094. [PubMed: 14593185]
28. Khan M, et al. Regulation of the probability of mouse odorant receptor gene choice. *Cell*. 2011; 147:907–921. [PubMed: 22078886]

29. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. *Genes & development*. 2012; 26:11–24. [PubMed: 22215806]
30. Edelman LB, Fraser P. Transcription factories: genetic programming in three dimensions. *Curr Opin Genet Dev*. 2012; 22:110–114. [PubMed: 22365496]
31. Clowney EJ, et al. Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell*. 2012; 151:724–737. [PubMed: 23141535]
32. Lomvardas S, et al. Interchromosomal interactions and olfactory receptor choice. *Cell*. 2006; 126:403–413. [PubMed: 16873069]
33. Montavon T, et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell*. 2011; 147:1132–1145. [PubMed: 22118467]
34. Kehayova P, et al. Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc Natl Acad Sci U S A*. 2011; 108:17195–17200. [PubMed: 21949399]
35. Nguyen MQ, et al. Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell*. 2007; 131:1009–1017. [PubMed: 18045541]
36. Magklara A, et al. An epigenetic signature for monoallelic olfactory receptor expression. *Cell*. 2011; 145:555–570. [PubMed: 21529909]
37. Clowney EJ, et al. High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Res*. 2011; 21:1249–1259. [PubMed: 21705439]
38. Lefebvre JL, et al. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature*. 2012; 488:517–521. [PubMed: 22842903]
39. Chess A. Monoallelic expression of protocadherin genes. *Nature genetics*. 2005; 37:120–121. [PubMed: 15678142]
40. Morishita H, Yagi T. Protocadherin family: diversity, structure, and function. *Curr Opin Cell Biol*. 2007; 19:584–592. [PubMed: 17936607]
41. Ribich S, et al. Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc Natl Acad Sci U S A*. 2006; 103:19719–19724. [PubMed: 17172445]
42. Monahan K, et al. Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression. *Proc Natl Acad Sci U S A*. 2012; 109:9125–9130. [PubMed: 22550178]
43. Guo Y, et al. CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proc Natl Acad Sci U S A*. 2012; 109:21081–21086. [PubMed: 23204437]
44. Kawaguchi M, et al. Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster. *J Biol Chem*. 2008; 283:12064–12075. [PubMed: 18204046]
45. Santoro SW, Dulac C. The activity-dependent histone variant H2BE modulates the life span of olfactory neurons. *eLife*. 2012; 1:e00070. [PubMed: 23240083]
46. Starr TK, et al. Positive and negative selection of T cells. *Annual review of immunology*. 2003; 21:139–176.
47. Monroe JG, et al. Positive and negative selection during B lymphocyte development. *Immunologic research*. 2003; 27:427–442. [PubMed: 12857986]
48. Li Q, et al. Locus control regions. *Blood*. 2002; 100:3077–3086. [PubMed: 12384402]
49. Feeney AJ. Epigenetic regulation of antigen receptor gene rearrangement. *Curr Opin Immunol*. 2011; 23:171–177. [PubMed: 21216580]

Highlights

- Stochasticity is important for the regulation of several mammalian gene families.
- It is often mediated via epigenetic silencing and nuclear architecture.
- Olfactory receptor expression is a prime paradigm of stochastic choice in neurons.
- Critical functions of nervous and immune systems are based on stochastic choices.

Figure 1. Allelic exclusion takes place in two phases where epigenetic mechanisms and nuclear repositioning play an important role

(a) In progenitor B cells both Ig κ alleles are initially DNA hypermethylated (red circle). At a later stage one allele becomes histone acetylated (green circle) and subsequently it undergoes demethylation, it while the other allele is recruited to heterochromatin. Finally, VJ recombination is initiated at the "active" allele leading to juxtaposition of the segments, while the "repressed" allele remains unrearranged [18]. (b) In the maintenance phase the rearranged (active) IgH allele remains in an euchromatic region, is marked by histone acetylation (green stars) and H3K4me3 (blue circles) and can undergo transcription (arrow) [49]. The unrearranged allele is recruited to pericentromeric heterochromatin, adopts a "closed" chromatin structure [19] and is probably marked by DNA hypermethylation (red stars) and repressive histone modifications (yellow circles) that prevent a secondary rearrangement.

Figure 2. Stochastic choice and stabilization of expression in OR genes

The OR genes are marked by H3K9me3 (in pink) and H4K20me3 (in purple) and they form specific nuclear foci. The repressive histone marks serve as docking sites for HP1 proteins that play an essential role in heterochromatin packaging and spreading. Each OR cluster may be regulated by a distant enhancer element (H-like element). A stochastically chosen OR allele moves out of the repressive nuclear focus into a permissive transcriptional environment, where it can interact with its cognate enhancer. Enzymatic complex(es) with histone demethylase and/or methylase activities carry out the epigenetic switch from H3K9me3/H4K20me3 to H3K4me3. Once a functional OR protein is produced, a feedback signal is initiated that prevents de-silencing of the rest of the OR genes probably through targeting a histone demethylase complex.

Figure 3. Genomic organization and stochastic promoter choice in the mouse *Pcdha* cluster

(a) The *Pcdha* gene cluster consists of 14 variable and 3 constant exons. The black blocks indicate identified enhancers. (b) The promoter of the variable exon α7 is stochastically chosen and activated, probably via its interaction with *cis*-regulatory sequences (HS5-1 and HS7), leading to the splicing of exon α7 to the set of constant exons and generation of the mature mRNA (c). (d) Most of the promoters of the variable exons and the enhancer HS5-1 are bound by complexes of CTCF/cohesin, while the c2 promoter and the HS7 enhancer are bound only by cohesin[42]. (e) The activated promoter α7 and the enhancers form a DNA-loop that is mediated by the CTCF/cohesin complexes. The rest of the variable promoters are getting marked by DNA methylation and histone modifications (H3K9me3 and H4K20me3), while they lose the CTCF/cohesin complexes, thus they are led to a repressive state.