



May the Odds Be Ever in Your Favor: Non-deterministic Mechanisms Diversifying Cell Surface Molecule Expression

Donnell L. Williams^{1,2†}, Veronica Maria Sikora^{1†}, Max A. Hammer^{1†}, Sayali Amin¹, Taama Brinjikji¹, Emily K. Brumley¹, Connor J. Burrows¹, Paola Michelle Carrillo¹, Kirin Cromer¹, Summer J. Edwards¹, Olivia Emri¹, Daniel Fergle¹, M. Jamal Jenkins^{1,2}, Krishangi Kaushik¹, Daniella D. Maydan¹, Wrenn Woodard¹ and E. Josephine Clowney^{2*}

OPEN ACCESS

Edited by:

Jin Xu,
Sun Yat-Sen University, China

Reviewed by:

Alexander Gimelbrant,
Harvard Medical School,
United States
Daniele Canzio,
University of California, San Francisco,
United States

*Correspondence:

E. Josephine Clowney
jclowney@umich.edu

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Developmental Epigenetics,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 05 June 2021

Accepted: 24 November 2021

Published: 11 January 2022

Citation:

Williams DL, Sikora VM, Hammer MA,
Amin S, Brinjikji T, Brumley EK,
Burrows CJ, Carrillo PM, Cromer K,
Edwards SJ, Emri O, Fergle D,
Jenkins MJ, Kaushik K, Maydan DD,
Woodard W and Clowney EJ (2022)
May the Odds Be Ever in Your Favor:
Non-deterministic Mechanisms
Diversifying Cell Surface
Molecule Expression.
Front. Cell Dev. Biol. 9:720798.
doi: 10.3389/fcell.2021.720798

¹MCDB 464 – Cellular Diversity in the Immune and Nervous Systems, University of Michigan, Ann Arbor, MI, United States,
²Department of Molecular, Cellular and Developmental Biology, The University of Michigan, Ann Arbor, MI, United States

How does the information in the genome program the functions of the wide variety of cells in the body? While the development of biological organisms appears to follow an explicit set of genomic instructions to generate the same outcome each time, many biological mechanisms harness molecular noise to produce variable outcomes. Non-deterministic variation is frequently observed in the diversification of cell surface molecules that give cells their functional properties, and is observed across eukaryotic clades, from single-celled protozoans to mammals. This is particularly evident in immune systems, where random recombination produces millions of antibodies from only a few genes; in nervous systems, where stochastic mechanisms vary the sensory receptors and synaptic matching molecules produced by different neurons; and in microbial antigenic variation. These systems employ overlapping molecular strategies including allelic exclusion, gene silencing by constitutive heterochromatin, targeted double-strand breaks, and competition for limiting enhancers. Here, we describe and compare five stochastic molecular mechanisms that produce variety in pathogen coat proteins and in the cell surface receptors of animal immune and neuronal cells, with an emphasis on the utility of non-deterministic variation.

Keywords: monogenic, monoallelic, stochastic gene choice, V(D)J recombination, Dscam, protocadherin, olfactory receptor, antigenic variation

INTRODUCTION

Despite the shocking complexity of eukaryotic life, eukaryotic genomes often contain less than 20,000 protein-coding genes. While most genes are expressed in a deterministic manner, a variety of molecular mechanisms have been discovered that expand the coding capacity of the genome by expressing cell surface molecules in a quasi-random manner. Expression systems that accomplish cell surface molecule diversification make use of genomic rearrangement, RNA splicing, and epigenetic restriction to create a vast array of molecular variants from a limited amount of DNA. In this way, the static information within genomes can generate a wider diversity of cells throughout the body or across unicellular populations.

TABLE 1 | Comparison of non-deterministic systems of cell surface molecule expression. For brevity, references are not included; they are provided throughout the main text description of each system.

	VSG	IgG	Dscam	Pcdh	ORs
Combinatorial diversification	Yes—construction of mosaic VSGs increases repertoire	Yes—V, D, and J exons are variably combined	Yes—exons 4, 6, and 9 are variably combined; multiple isoforms per cell	Yes—cells can express isoforms from A, B, and G clusters; some cells express multiple genes from one cluster	No
Monoallelic	N/A—active VSG copied from a “genomic archive;” expression sites can be hemizygous	Yes	No	Sometimes	Yes
Exclusive (i.e. exactly one isoform/cell)	Yes	Yes	No	Cells generally express isoforms from A, B, and G clusters. The choice within cluster is not necessarily exclusive	Yes
Dependence on limiting enhancer	Active VSG associates with genomic locus encoding splice-leader RNA	Yes, for proximal V promoter selection	At the RNA level, dependent on unique RNA “chooser” elements	Yes, e.g. HS5-1 for PCDHA	Yes, Greek Islands
Mechanism of choice/variation	Recombination into active site, active site switching, construction of mosaic VSGs	Recombination and AID-induced point mutation	Alternative splicing	Promoter choice via limiting enhancer(s)	Promoter choice via limiting enhancers
Expression choice in each cell is initially more promiscuous, and then refines	Yes	No	N/A, not exclusive	Yes	Yes
Choice is stable once refined	Choice is heritable. Switching is critical, but unclear how it is induced	Yes	No	Unknown	Yes
Feedback	Selection by immune system clearance	Unfolded protein response	Unknown	Unknown	Unfolded protein response
Function of non-deterministic choice	Immune system can't predict what antigen will be expressed next, mosaic VSGs expand repertoire	Pathogen can't predict what antibodies will be present, allows defense against novel pathogens that were not predicted by evolution	Allows neurons of the same ontogenetic type to have distinct barcodes, and allows neurons to respond differently to themselves than to ontogenetically identical sisters	Allows neurons of the same ontogenetic type to have distinct barcodes, and allows neurons to respond differently to themselves than to ontogenetically identical sisters	Provides a concise mechanism for activating OR expression; new ORs can be expressed without evolution of new transcriptional mechanism; only need one regulatory system instead of 1,000
Drawbacks of non-deterministic choice	Not obvious	Since antibodies are produced randomly, many arise that bind self-antigens. These must be selected against	Not obvious	Not obvious	Requires receptor-dependent mechanisms to wire OSNs to olfactory bulb glomeruli
Function of restricted expression and diverse cell surface phenotypes	Prolongs infection by allowing host immune system to “see” only one VSG at once	Allows binding of diverse and novel antigens; compartmentalization allows cellular somatic selection of effective receptors	Neuronal self-identification and self-avoidance	Neuronal self-identification and self-avoidance	Olfactory perception—each cell senses limited and distinct odorants

This diversity is crucial for the proper functioning of many different biological systems. The immune system, for example, relies on diverse antigen receptors to bind to and recognize an incredible range of potential pathogens and harmful molecules. Without the stochastic mechanisms driving this variation in expression, the proper functioning of the immune system would be severely compromised. The nervous system is similar

in that it also relies on diversity in gene expression for proper functioning. Like the immune system, neurons in chemosensory systems express diverse receptors to bind a wide array of environmental molecules. In addition, neurons, even of the same class, must be sufficiently different from one another in order to properly identify self vs. non-self. Finally, pathogens also diversify their surface molecules in an arms race with the adaptive

immune system. Thus, understanding the mechanisms that produce non-deterministic cellular heterogeneity is an important area of study.

In this review, we will focus specifically on non-deterministic processes that select one or a few surface molecules to be expressed on a particular cell from among many copies of similar sequences encoded in the genome. We highlight 5 such systems: the expression of variable surface glycoproteins (VSGs) by the parasite *Trypanosoma brucei*, pathogen identification by B cell and T cell receptors, neuronal self-avoidance through expression of Dscams and protocadherins, and the perception of stimuli through the olfactory system. While other reviews have compared subsets of these systems, here we broaden the scope of the comparison by considering both single-celled organisms and animals and by considering both neuronal and barrier functions (Magklara and Lomvardas 2013; Khamlichi and Feil 2018; Aresta-Branco et al., 2019a). In addition to comparing molecular mechanisms, we highlight the distinct types of utility gained by non-deterministic expression in different systems. Often, procedural or algorithmic mechanisms are simply more concise than deterministic mechanisms. In other cases, unpredictability in molecular outcomes is itself crucial for cellular function.

There are many similarities across these five examples (summarized in **Table 1**). First, they all have some type of restriction mechanism, often heterochromatin-based, that ensures that all of the coding sequences that could possibly be expressed aren't expressed at the same time. Each system also involves stochastic selection of a single (or a few) isoform(s) that will be expressed. For antigen receptors, Dscams, protocadherins, and olfactory receptors, stochastic selection involves a unique enhancer or locus control region. Such a region or enhancer has not yet been identified for VSGs. Lastly, in three of these systems, there are feedback mechanisms downstream of selection that can act to help correct any flaws that were made during selection. In antigenic (VSG) variation, this feedback is whether or not the cell survives the host immune system. In V(D)J recombination, feedback takes place within the germinal center when higher affinity B cells win the competition for antigen. In olfactory receptor choice, the feedback mechanism allows the cell to choose a different olfactory receptor gene if it initially chose a flawed one—or stops the cell from choosing another gene if the one it already chose is functional. Although similar feedback processes may take place in Dscam and protocadherin expression, they have not yet been discovered.

While we restrict our analysis here to mechanisms that diversify cell surface molecule repertoires by choosing among genetically encoded paralogues, we note that all biological diversification ultimately relies on noise in genome replication that produces mutations, and that noise is often harnessed and regulated to do biological work. For example, HIV immune evasion has been suggested to result from the virus's retention of an unusually error-prone replication enzyme, and switches between lytic and latent phases are thought to occur stochastically (Roberts et al., 1988; Weinberger and Weinberger 2013; Cuevas et al., 2015). Behavioral switches are also likely governed by probabilistic rather than deterministic mechanisms. The degree of variation in gene expression between cells is itself subject to selection, and such variation can alter the penetrance of mutant alleles (Raj et al.,

2010; Metzger et al., 2015; Duveau et al., 2018) Stochastic processes can also reduce the fitness costs of mutations, as in the case of X inactivation in female mammals. While the processes of life contravene entropy, in many cases the otherwise robust and predictable mechanisms of cellular development allow molecular noise to peek through in a regulated manner to influence phenotype.

As we discuss throughout, the monogenic and/or monoallelic expression of cell surface molecules allow each of these systems to appropriately interact with the outside or extracellular world. The functional purpose of selecting cell surface molecules in a non-deterministic rather than a predictable manner varies across them. In some cases, non-deterministic processes may be the only way for cells in otherwise almost identical environments and with identical differentiation regimens to become distinct from one another. In the nervous system, for example, groups of neurons that are developmentally equivalent and located in the same location can produce different cell-surface proteins by randomly selecting and expressing certain gene segments or genes—as is the case with olfactory receptors in olfactory sensory neurons. Non-deterministic expression systems likely also allow for a larger array of different proteins to be made than can otherwise be deterministically encoded by the genome, as is likely the case for the immune system. Because it is inherently unpredictable, non-deterministic expression may also increase fitness for hosts and pathogens locked in battle with one another.

Definitions

In these fields, the terms “stochastic” and “random” are used to refer to processes in which knowing the ontogenetic identity of a cell predicts a distribution of possible gene expression choice but is insufficient to deterministically predict cell surface molecule expression. We note that in most of these systems, molecular choices follow biased distributions—for example, olfactory receptor choice is biased by position in the olfactory epithelium, VSG choice by the time course of infection, and Dscam choice by the neuronal cell type. Biased distributions are consistent with the mathematical definitions of stochastic or random, and we continue to use those terms here. We use the terms “non-deterministic,” “probabilistic,” and “unpredictable” as additional descriptors. Further, we use “monoallelic” to refer to molecular choice between two copies of the same gene, and “monogenic” to refer to selection among paralogues. We include gene families here, for example the Dscams, in which surface molecule choice is not exclusively monogenic, i.e. where multiple choices are made in each cell but most of the available choices are still repressed. We note that in the VSG field, expression of a single VSG paralogue per trypanosome is typically referred to as “monoallelic” expression. For consistency with the other topics covered here, we use the term “monogenic.”

VARIANT SURFACE GLYCOPROTEINS

Many pathogens have learned to survive in host environments that are hostile to their growth. One such method that pathogens have evolved is antigenic variation. Here, we will discuss coat protein switching in trypanosome infection as a model. *Trypanosoma brucei* is a single-celled eukaryotic pathogen that

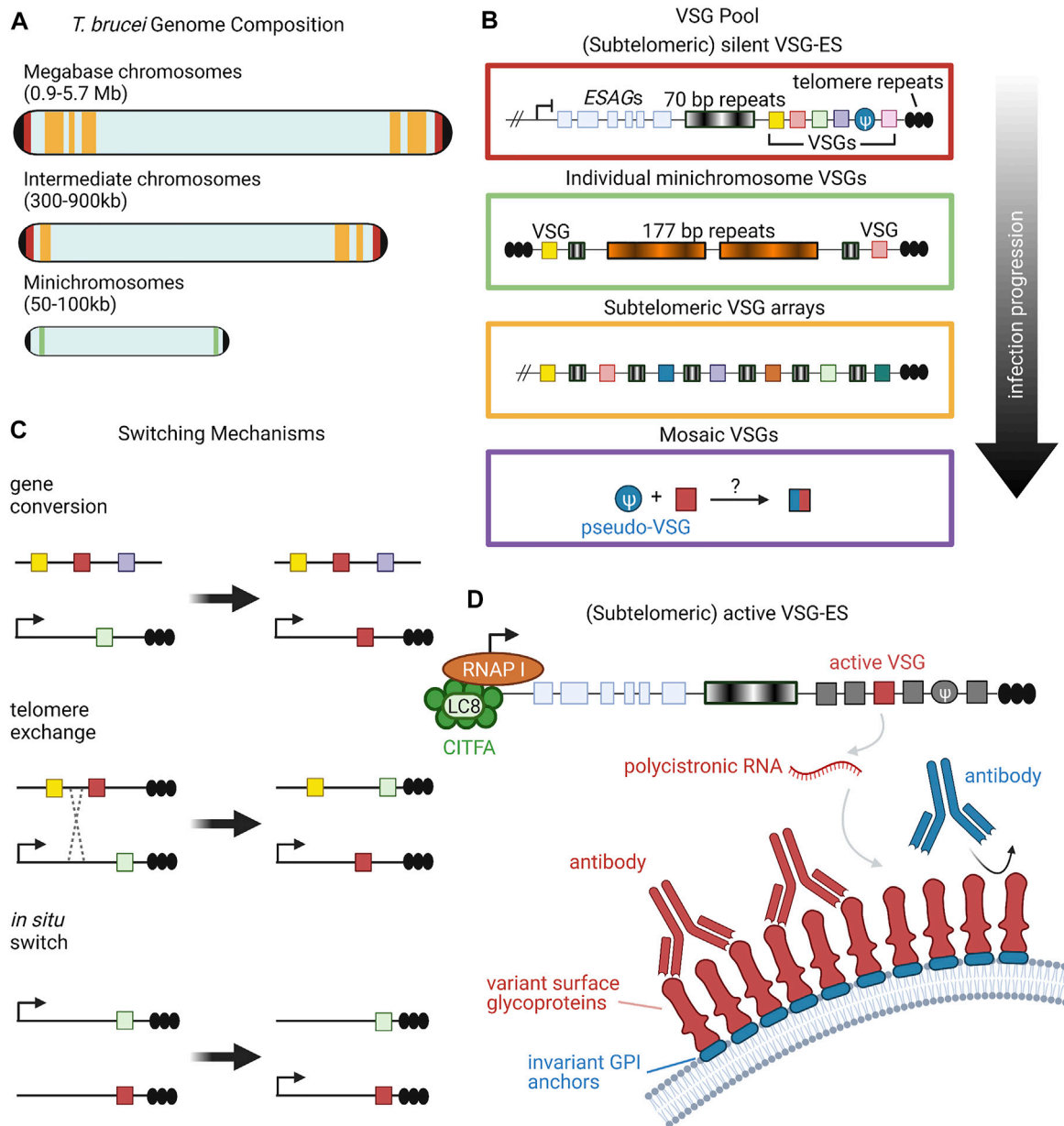


FIGURE 1 | Organization and Expression of VSGs. **(A)** *T. brucei* has an unusual karyotype consisting of large, megabase-sized chromosomes, intermediate chromosomes, and minichromosomes. Size ranges of each chromosome type are listed in parentheses (Berriman et al., 2005). Black caps on the ends of the chromosomes represent telomeres. Red, yellow, and green bands denote the typical locations of different VSG repertoires, corresponding to the colored-coded insets in panel **(B)**. **(B)** Candidate VSG genes are located on megabase and intermediate chromosomes within silent subtelomeric bloodstream expression sites (VSG-ES; red box) or subtelomeric arrays (yellow box). Individual VSG genes can also be found in subtelomeric regions of minichromosomes (green box) or can be generated from recombination of intact and/or VSG pseudogenes from various sources (purple box). Throughout the course of infection, *T. brucei* will draw upon the VSG pool in a semi-predictable manner according to the location of candidate genes; for example, VSG gene arrays from silent expression sites are typically used before minichromosome VSGs (Sima et al., 2019). **(C)** VSG expression proceeds from a single active expression site. To shift expression to a new VSG gene, *T. brucei* can employ one of three switching mechanisms: gene conversion, telomere exchange, or *in situ* switching (Liu et al., 1983; Rudenko et al., 1996; Horn and Cross, 1997; Robinson et al., 1999; Li, 2015). **(D)** RNA polymerase I transcribes polycistronic RNAs from active VSG expression sites. The CITFA transcription factor complex, which in *T. brucei* consists of CITFA subunits 1–7 (green circles) complexed with LC8/DYNLL1 (light green oval), is a basal transcription factor required for RNA pol I initiation (Kirkham et al., 2016). The active VSG gene is typically transcribed last, preceded by expression-site associated genes (ESAGs) (Pays et al., 2001; Hertz-Fowler et al., 2008). VSG RNA is translated into variant surface glycoproteins, which form a densely-packed coat that prevents recognition of underlying cell-surface molecules, such as GPI anchors (Hertz-Fowler et al., 2008; Horn and McCulloch, 2010). Figure inspiration was drawn from various sources (Rudenko et al., 1996; Berriman et al., 2005; Li, 2015). All figures in this review were created using BioRender.

has dedicated a large amount of its genome to this process. This parasite is found mostly in sub-Saharan Africa and is the cause of a vector-borne disease known as sleeping sickness. *T. brucei* is coated by a dense layer of variant surface glycoproteins (VSGs) and is able to switch which VSG is expressed in order to evade host immune systems (Boothroyd et al., 2009; Hoeijmakers et al., 1980; Sima et al., 2019). The high density of surface VSG molecules shields the pathogen's other non-variable surface proteins (**Figure 1**), making the pathogen's immunological identity tied to the particular VSG it expresses (Hertz-Fowler et al., 2008; Horn and McCulloch, 2010).

Research suggests that there are about 2000 genes that constitute the VSG repertoire of *T. brucei*, clustered into subtelomeric arrays as well as on several minichromosomes (**Figure 1A**) (Cross et al., 2014). Though each individual *T. brucei* organism expresses only a single VSG gene at a time, VSG switching has been shown to occur at a rate as high as 10^{-3} switches per cell per generation (Mugnier et al., 2015; Turner and Barry, 1989). This high switching rate, along with the large number and diversity of cells present, leads to a sinusoidal pattern of infection where the immune system eliminates cells expressing a given VSG, but not before new variants arise in the population. These variants then grow in number, only to be wiped out again by the immune system, followed by the emergence of new variants (Mugnier et al., 2016). This constant back-and-forth between new VSG variants and the host immune system allows for *T. brucei* to remain inside of a host for long periods of time, creating chronic infections. Interestingly, studies of *T. brucei* population dynamics have begun to reveal semi-predictable patterns in VSG expression based on gene location and other gene family characteristics (**Figure 1B**), but much is still unknown about the level of determinism in the system (Morrison et al., 2005; Mugnier et al., 2015).

Multiple overlapping mechanisms accomplish this dynamic VSG switching (**Figure 1C**). The predominant mechanism is duplicative VSG gene conversion, in which a silent VSG gene is copied into an active expression site (Liu et al., 1983; Robinson et al., 1999; Li, 2015). VSG expression can also swap via *in situ* switching where a previously silenced expression site is activated, while the previously active site is silenced (Horn and Cross, 1997). A third mechanism is telomere exchange, where telomeric regions undergo crossing over that swaps which VSG is downstream of the active promoter (Rudenko et al., 1996). The field has primarily focused on these first two mechanisms, with telomere exchange still relatively underexplored, so we will focus more on gene conversion and *in situ* switching in this review.

As in all the monogenic expression systems described in this review, repression of the majority of possible loci is a necessary condition for restricted use of the chosen locus. The sub-telomeric location of VSG expression sites plays a part in their repression (Ersfeld et al., 1999; Berriman et al., 2005). Telomere proximity is inversely related to transcriptional activity of genes generally (Robin et al., 2014), and this trend holds true for DNA Pol I transcribed genes such as VSG genes (Glover and Horn, 2006). The telomere binding protein RAPI is an essential component of the telomere complex and has been associated with VSG repression (Yang et al., 2009). The protein phosphatidylinositol 5-phosphatase binds to RAPI, and, together with phosphatidylinositol 5-kinase, helps to control VSG gene

repression near the telomere by phosphorylating and dephosphorylating key regulatory molecules (Cestari and Stuart, 2015).

VSG Expression Sites

The trypanosome genome has 20–40 polycistronic, sub-telomeric expression sites (ES) that promote transcription of VSGs as well as ES-associated genes (*ESAGs*) (Hertz-Fowler et al., 2008; Pays et al., 2001). They contain a Pol I promoter and are typically around 45 kb in length, with the VSG gene the most distal gene transcribed (**Figure 1D**) (Pays et al., 2001; Hertz-Fowler et al., 2008). We will focus our attention on expression sites active during the bloodstream stage of expression (bloodstream expression site, BES). Interestingly, though VSG proteins are monogenically expressed, it has been observed that multiple BES can be transcriptionally active at a time (Kassem et al., 2014). The additional BES transcripts do not fully elongate, are transcribed at lower levels, and are not translated, indicating additional regulation at the transcriptional and post-transcriptional levels to maintain monogenic expression (Kassem et al., 2014). Due to differing recombination into BESs, the two alleles of a particular BES could contain different contents; regardless, expression is from only one BES per cell and is therefore monoallelic.

VSG Induction, Inheritance, and Switching

Monogenic expression and switching of a single VSG gene is what allows the parasite to successfully evade the host immune system. In fact, parasites that express multiple VSG proteins at once are quickly cleared by the immune system (Aresta-Branco et al., 2019b). VSG expression initiates in parasites that reside in the salivary gland of the tsetse fly, prior to bloodstream infection. Recent data has shown that multiple VSG genes are initially transcribed within pre-metacyclic cells, with a single gene being expressed within mature metacyclic cells (Hutchinson et al., 2021). A “race” model has been proposed to explain this phenomenon in which different VSG expression sites race to hit a certain threshold level of transcription. Once a particular gene hits this threshold, the other transcribed expression sites become downregulated, possibly due to the limited expression machinery being used up at this single site, or by the actively transcribed RNA transcripts silencing expression at the other sites (Hutchinson et al., 2021). The particular transcribed VSG and its localization to the nuclear expression site can be inherited following cell division and this inheritance depends on the chromatin assembly factor CAF1 (Faria et al., 2019). Remarkably, simply loosening chromatin structure through ectopic overexpression of the high-mobility group box protein TDP1 is sufficient to allow expression of multiple VSGs per cell (Aresta-Branco et al., 2019b).

While the choice of active VSG can be stable within the life of a cell and through cell division, occasional VSG switching is critical for immune evasion and long-term infection. How VSG switching is regulated—whether this is a probabilistic event or induced by parasite or host factors—remains unknown. The molecular mechanisms that induce VSG exchange are also mysterious. Some possible explanations include collapse of the replication fork due to continuous VSG transcription (Glover et al., 2013), or translocations triggered by frequent DNA damage,

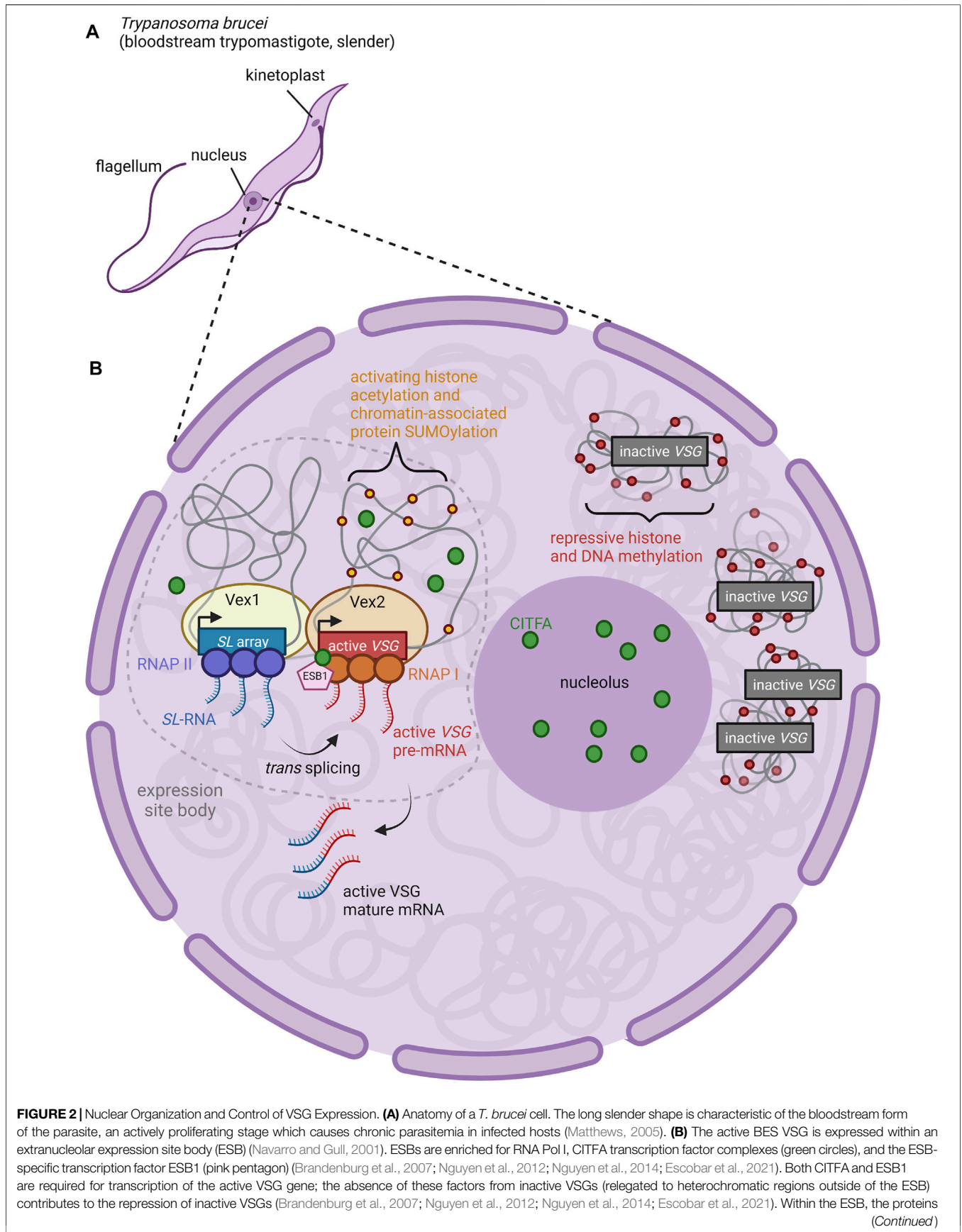


FIGURE 2 | Nuclear Organization and Control of VSG Expression. (A) Anatomy of a *T. brucei* cell. The long slender shape is characteristic of the bloodstream form of the parasite, an actively proliferating stage which causes chronic parasitemia in infected hosts (Matthews, 2005). **(B)** The active BES VSG is expressed within an extranucleolar expression site body (ESB) (Navarro and Gull, 2001). ESBs are enriched for RNA Pol I, CITFA transcription factor complexes (green circles), and the ESB-specific transcription factor ESB1 (pink pentagon) (Brandenburg et al., 2007; Nguyen et al., 2012; Nguyen et al., 2014; Escobar et al., 2021). Both CITFA and ESB1 are required for transcription of the active VSG gene; the absence of these factors from inactive VSGs (relegated to heterochromatic regions outside of the ESB) contributes to the repression of inactive VSGs (Brandenburg et al., 2007; Nguyen et al., 2012; Nguyen et al., 2014; Escobar et al., 2021). Within the ESB, the proteins

(Continued)

FIGURE 2 | VEX1 (yellow oval) and VEX2 (light orange oval) complex together, associating an mRNA splicing locus (SL array, blue rectangle; RNAP II, purple circles) with the active VSG expression locus (active VSG gene, red rectangle; RNAP I, orange circles) (Glover et al., 2016; Faria et al., 2021). Thus, transcription of the splice leader (SL) RNA by RNAP II proceeds adjacent to transcription of the active VSG pre-mRNA by RNAP I. The 5' end of the SL RNA is then spliced in trans to the VSG pre-mRNA to form mature VSG mRNA (Faria et al., 2021). Figure inspiration was drawn from various sources (López-Farfán et al., 2014; Martínez-Calvillo et al., 2019; Nguyen et al., 2014; Escobar et al., 2021; Faria et al., 2021).

such as double stranded breaks, within unstable regions surrounding expression sites (Boothroyd et al., 2009). However, loss of RECQ2, a helicase which repairs DNA breaks within the telomere, leads to an increase in DNA recombination, indicating a possibility that double stranded breaks are not responsible for inducing VSG switching (Devlin et al., 2016). In contrast, VSG transcription and DNA replication have been shown to be associated with one another (Devlin et al., 2016). Thus, VSG switching could be induced by DNA fragility brought about by DNA replication (Devlin et al., 2016).

Telomere length, telomere stability, and the regulation of the chromatin structure surrounding VSG expression sites has also been shown to be important for VSG switching (Hovel-Miner et al., 2012; Aresta-Branco et al., 2016). The degree to which this VSG expression choice is stochastic versus deterministic has also come into question with studies revealing a degree of predictability in VSG emergence, which could result from either ordering of VSG choice or from differential selection (Morrison et al., 2005; Mugnier et al., 2015).

Expression Site Activation and In Situ Switching

While only one BES at a time produces an actively translated product, the active BES can switch between the repertoire of available BES through *in situ* switching (Figure 1C). Several factors have been identified as characteristic features of the active BES that must be altered in order for *in situ* switching to occur (Cestari and Stuart, 2018). Reminiscent of the importance of nuclear organization in *OR* gene selection (described below), active BES are localized to an extranuclear region termed the expression site body (Figure 2) (Navarro and Gull, 2001). The expression site body and active BES promoter are enriched for Pol I along with the *basal class I transcription factor A* (CITFA) complex (Brandenburg et al., 2007; Nguyen et al., 2012; Nguyen et al., 2014). The novel transcription regulator NLP similarly associates selectively with the active BES (Narayanan et al., 2011).

What mechanisms ensure that only a single BES can produce functional VSG in each cell? Work in the last 5 years has focused on two factors identified in genetic screens that are required for VSG expression: *VSG exclusion 1* (VEX1) and *VSG exclusion 2* (VEX2). VEX1 has been shown to positively regulate the active VSG site in *cis* while also negatively regulating all other VSGs in *trans* (Glover et al., 2016). VEX1 binds to VEX2 independently of transcription, and together, they are responsible for VSG exclusion (Faria et al., 2019). Recent work suggests that the single VSG chosen for expression in a particular cell is physically associated with the genomic locus that encodes mRNA splice-leader sequences

(Faria et al., 2021). In *T. brucei*, mature mRNA is produced by *trans*-splicing to leader sequences produced from this locus. The active BES can associate with the splice-leader locus in *trans*, across chromosomes (Faria et al., 2021). VEX1 associates with the splice leader locus, and VEX2 with the actively expressed VSG in the BES.

Thus, VSG transcription and mRNA splicing take place within a specific compartment of the nucleus and are associated closely with VEX1 and VEX2 proteins (Figure 2) (Faria et al., 2021). The coalescence of the VEX1-bound splice leader locus and the VEX2-bound BES into the expression site body may serve to activate Pol I transcription and subsequently induce repression of the remaining BESs (Glover et al., 2016; Schulz and Papavasiliou, 2016; Cestari and Stuart, 2018; Faria et al., 2021). Transcription-mediated gene silencing is a hallmark of heterochromatinization in other stochastic systems such as yeast mating-type switching, and could be involved in olfactory receptor heterochromatinization as well (Allshire and Madhani, 2018). Another possibility is that failure to concentrate access to the splice leader cassette and transcription factors on one VSG gene allows multiple VSGs to be expressed at lower levels.

Chromatin modifications also appear to play an important role in BES activation. The active BES is significantly depleted of histones, especially H3, in comparison to other, silent BESs (Stanne and Rudenko, 2010). In support of this, the knockdown of H1, H3, H3.V, and H4.V increased the accessibility and transcription at previously inactive BES promoters and VSG genes (Povelones et al., 2012; Reynolds et al., 2016; Schulz et al., 2016; Müller et al., 2018). This effect has been shown to be mediated by chromatin remodeling proteins such as ASF1A, CAF-1b, and SIR2rp1, alongside a handful of histone methyltransferases and acetylases/deacetylases (Alsford et al., 2007; Figueiredo et al., 2008; Kawahara et al., 2008; Wang et al., 2010; Alsford and Horn, 2012). SUMOylation also plays a crucial role, with SUMOylated chromatin-associated proteins serving as a distinct marker of the active BES in the expression site body (López-Farfán et al., 2014). The VSG transcriptional activator SNF2PH is recruited to SUMOylation-rich BES where it is itself SUMOylated to subsequently facilitate Pol I transcription (Saura et al., 2019). Pol I is then further regulated by activating SUMOylation via TbsIZ1/PIAS1 (López-Farfán et al., 2014). All of these SUMOylation events appear to be localized to the active BES.

While nuclear localization, transcription factor recruitment, and chromatin modification have all been shown to be relevant in VSG expression at the selected BES, the order and dominance of these activating events is still unclear. It is also still uncertain

what induces these activation signals to switch between BESs during *in situ* switching.

Duplicative Gene Conversion

Early genetic experiments revealed that VSG switching can involve the overwriting of genomic loci (Horn, 2014). A form of recombination, duplicative gene conversion involves the removal of the active VSG gene from the expression site, which is replaced by a duplicated form of a previously silent VSG gene (**Figure 1C**) (Liu et al., 1983; Robinson et al., 1999). Boothroyd et al. found that gene conversion switches are initiated by DNA double-strand breaks which are subsequently repaired by homologous recombination (Boothroyd et al., 2009). Each VSG gene possesses an upstream region of 70 bp repeats. Double-strand breaks adjacent to these repeats were both necessary and sufficient to induce VSG switching, suggesting that the repeats serve as a guide for homologous recombination that allows for the active VSG site to be overwritten (Boothroyd et al., 2009). The BESs, along with many of the silent VSG arrays, are located in sub-telomeric regions of the genome (Cross et al., 2014). These regions are inherently unstable portions of the genome where recombination and double-strand breaks frequently occur (Glover et al., 2013; Horn, 2014). As such, it is suspected that *T. brucei* takes advantage of this natural instability to induce VSG gene conversion; however, alternative mechanisms for DNA lesion production have been proposed (reviewed in da Silva et al., 2018). VSG recombination requires RAD51 and BRCA2, while TOPO3 α has been shown to suppress recombination and restrict it to the 70bp repeats in partnership with RMI1 (McCulloch and Barry, 1999; Hartley and McCulloch, 2008; Kim and Cross, 2010; Kim and Cross, 2011).

Mosaic VSGs

While trypanosomes predominantly switch between existing intact VSG genes at the beginning of an infection, long read sequencing has confirmed that there is a significant increase in the number of novel mosaic VSG genes as infection time increases (Jayaraman et al., 2019; Mugnier et al., 2015). Because 80% of *T. brucei*'s ~ 2000 VSG genes are incomplete or pseudogenes, the repertoire of complete genes is eventually exhausted during chronic infections (Berriman et al., 2005; Cross et al., 2014). Once the majority of complete genes have been expressed and recognized by the immune system, *T. brucei* utilizes segmental gene conversion to merge fragments of different VSG genes in what is termed mosaic recombination (**Figure 1B**) (Mugnier et al., 2015). This process is not uncommon, as other pathogens have been known to utilize segmental gene conversion to further diversify their pool of variant genes (Zhuang et al., 2007). Trypanosomes are able to construct functional mosaic VSG genes from pseudogenes and gene fragments, suggesting that *T. brucei*'s large abundance of partial genes are important for continued diversification (Hall et al., 2013). It is still unclear what cellular process is used to

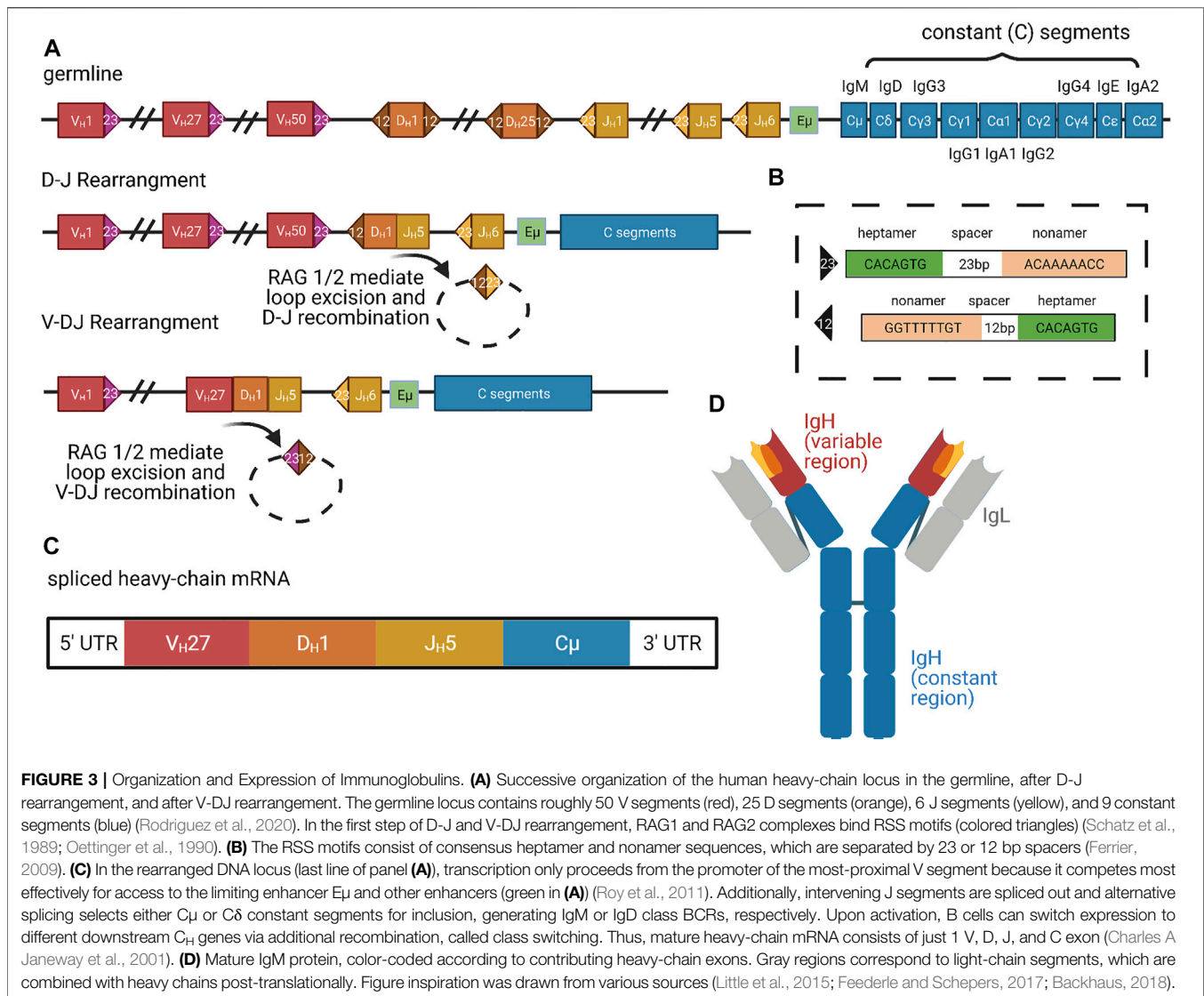
merge the VSG segments together. One possibility is that mosaics are generated by homologous recombination within the VSG gene similar to duplicative recombination or by crossover events, as in telomeric exchange. It is also unknown whether mosaic formation occurs within expression sites, or if instead they are formed somewhere else in the genome before being moved into the expression site.

Sleeping sickness remains a deadly and difficult to treat disease, so increasing our understanding of the mechanisms that allow this parasite to evade host immune systems will provide advances in our ability to fight *T. brucei* infections. A more detailed analysis of remaining questions in the field is reviewed by McCulloch and colleagues (McCulloch et al., 2017). Similar methods of variation utilized by trypanosomes can also be found in the systems they are meant to evade. Just as antigen diversity aids pathogens in evading the immune system, antigen receptor diversity allows for greater detection. B and T cells in the immune system create this diversity through stochastic genome editing. This process is often initiated by the introduction and subsequent repair of DNA double-stranded breaks, similar to gene conversion in VSG's (Papavasiliou and Schatz, 2000). It has also been suggested that B and T cells expand their receptor diversity via segmental gene conversion, similar to mosaic VSGs (Barbet and Kamper, 1993).

ANTIGEN RECEPTOR DIVERSITY

The coexistence of host and pathogen has largely driven the diversification of both the host's immune surveillance and the pathogen's antigenic determinants (Chang et al., 2011). Mammalian genomes contain approximately 20,000 protein-coding genes, and yet the B and T cells of the adaptive immune system produce receptors that can bind to a vast array of arbitrary antigens regardless of evolutionary experience. Receptor-level diversity was ultimately shown to be produced via two stochastic processes: V(D)J recombination, which alone can generate 10^{11} possible binding domains, and somatic hypermutation, which can introduce mutations in any of these recombination products to further expand receptor possibilities (Janeway et al., 2001). These processes allow for a truly outstanding level of diversity to emerge from just a few germline-level genes, preparing the immune system for any antigen it might face without taking up very much genomic space. Much as learning mechanisms in the nervous system allow animals to relate arbitrary sensory stimuli to the contexts in which they are experienced, selective processes during B and T cell development in the context of an immune response shape cellular immune responses according to the "meaning" of self, benign, or pernicious antigens.

Antibodies, which are immunoglobulin proteins, possess variable binding surfaces that can recognize diverse antigens. These antibodies can be secreted in the serum or bound to the surface of B lymphocytes to form B cell receptors (BCRs). T lymphocytes also have surface receptors (TCRs) that recognize antigens in combination with antigen presenting



major histocompatibility complex (MHC) proteins. Like TCRs and BCRs, MHC proteins are present in the population as diverse alleles. While TCR and BCR diversity is generated via somatic mechanisms, population-level MHC diversity is maintained at the germline level via balancing selection.

Immunoglobulin Gene Loci and V(D)J Recombination

Immunoglobulins are composed of covalently-linked heavy and light chains, both of which possess a variable N-terminus that recognizes antigens and a constant C-terminus that can recruit effectors (Schroeder and Cavacini, 2010). Here, we will focus on the generation of BCRs and antibodies as a model for immunoglobulin diversification. The germline-encoded heavy chain locus produces IgM and IgD constant regions via

alternative splicing; both IgM and IgD can be membrane-bound or secreted as antibodies. DNA rearrangements of the heavy chain locus during the course of the lymphatic germinal center reaction can also produce secreted IgG, IgA, and IgE antibodies by joining the variable N-terminus to different constant regions; we will focus on variable region diversification. V(D)J recombination during lymphocyte development produces the initial diversity of mature IgM and IgD. Once B cells bind their given antigen, somatic hypermutation in the germinal centers allows for further diversification of the variable region alongside the production of IgG, IgA, and IgE through class switching.

In V(D)J recombination, the N termini of the heavy and light chains are rearranged to bring distinct V (variable) segments, followed by the D (diversity) segments, and then the J (joining) segments into proximity with the constant region (Figure 3) (Schroeder and Cavacini, 2010). The human heavy chain locus on

chromosome 14 possesses roughly 50 functional V segments, 27 functional D segments, and six functional J segments (Rodriguez et al., 2020). The two major classes of light chains are kappa and lambda, both of which do not possess D segments but still undergo VJ recombination. The kappa locus is on chromosome 2 with roughly 44 functional V segments and 5 J segments, whereas the lambda locus is on chromosome 22 with roughly 37 functional V segments and only 1 J segment (Collins and Watson, 2018; Watson et al., 2015). The ability to create combinatorial V(D)J regions allows for an incredible diversity of heavy and light chains, which are both combined to further expand the possibilities for mature immunoglobulin proteins. There are roughly 3.5×10^6 potential combinatorial products that can arise from these V(D)J and heavy-light chain pairings, and final protein products are additionally varied by junctional diversification that occurs during recombination. Recombination is induced by RAG1/2 (recombination-activating gene) which target discrete locations within the immunoglobulin loci through conserved and repeated DNA sequence elements (**Figure 3A**) (Schatz et al., 1989; Oettinger et al., 1990). As the V(D)J recombination process is “settled science,” we refer readers to other reviews or textbooks for more detailed description.

Remarkably, diversification of antigen binding repertoires through alteration of germline DNA has evolved more than once. In the lamprey, leucine rich repeat (LRR) proteins are diversified during lymphocyte maturation via insertion of LRR modules from flanking regions of the locus (reviewed in Boehm et al., 2012). Random combinatorial usage of immunoglobulin modules has also arisen in other systems: neuronal self-recognition in insects is mediated by randomized alternative splicing of the *Dscam* immunoglobulin locus.

Enhancer-Mediated Restriction

While V(D)J recombination removes V regions proximal to D or J segments, distal V options remain intact, and each V has its own upstream promoter. Nevertheless, transcription always begins at the V region most proximal to D/J and therefore ensures that only a single V—the most proximal—is included in the transcript (Roy et al., 2011). This selection was initially thought to be performed by a limiting enhancer element located between the V and D/J domains, called E μ (**Figure 3A**) (Serwe and Sablitzky, 1993; Li and Eckhardt, 2009). More recent work has suggested that several additional enhancers, including 3'RR and DICE, participate in a complex promoter selection process (Bébin et al., 2010; Roy et al., 2011). The most proximal V region promoter that remains after recombination is thought to compete most effectively for looping interactions with the limiting enhancers, thus conferring deterministic use of the most proximal V in the context of stochastic removal of alternate distal Vs. As we will describe below, the process of clustered protocadherin transcription in mammalian neuronal self-recognition also involves competition among nearly identical promoters for access to a single enhancer, and expression variability is produced by suppression of spatial bias for the proximal promoter, rather than recombination of different segment choices into proximity with the enhancer.

Monoallelic Expression

Similar to the selective expression of a single allele in the OR, PCDH, and VSG systems, there is extensive evidence that each B cell expresses only a single BCR, which makes each B cell specific for one particular antigen (Weiler, 1965; Vettermann and Schlissel, 2010). This specificity is important for subsequent clonal selection of antibody-producing cells and proper immune response. Interestingly, the *Ig* alleles are transcribed biallelically early in B cell development, indicating that transcriptional activation alone does not govern the allelic exclusion of these loci (Singh et al., 2003). At the level of translation, transcripts from *Ig* genes that have not undergone proper V(D)J recombination possess premature stop codons that prevent production of functional protein (Bühler et al., 2004; Eberle et al., 2009). Moreover, B cells co-opt the unfolded protein response to trigger differentiation in response to BCR translation (Hetz et al., 2020). A similar process links olfactory receptor choice, stabilization of singular olfactory receptor translation, and olfactory neuron differentiation (Dalton et al., 2013).

Studies have suggested that complete V(D)J recombination of one allele induces the suppression of the second non-recombined allele, preventing subsequent recombination and productive transcription (Vettermann and Schlissel, 2010). This is supported by the observation that D-J recombination occurs in both IgH alleles, yet only one productive V-DJ recombination proceeds (Jung et al., 2006). To accomplish this, the active recombination of the locus appears to induce RAG- and ATM-mediated repositioning of the inactive allele to repressive heterochromatin alongside inducing locus decontraction that has been associated with recombination inhibition (Goldmit et al., 2005; Hewitt et al., 2009). Furthermore, the production of a complete immunoglobulin protein chain then induces progression of B cell development that subsequently downregulates RAG proteins to prevent further recombination (Grawunder et al., 1995; Galler et al., 2004). This model allows for developing B cells to make multiple attempts at performing proper recombination, as complete suppression of the alternative allele does not occur until one of the alleles has produced protein. However, in order for this process to produce a monoallelic product the induction of recombination must be asynchronous.

Early models suggested that the low rate of recombination allowed for a probabilistic first-come, first-serve mechanism where allelic selection was purely based on which allele happened to recombine first (Perry et al., 1980; Liang et al., 2004), but continued studies have revealed that the process is likely more controlled than this. It has been shown that the selected allele is replicated first and localized to the euchromatic nuclear center whereas the non-selected allele is found in the repressive heterochromatin of the nuclear periphery (Mostoslavsky et al., 2001; Skok et al., 2001). The active allele is subsequently found to have activation signatures: hypomethylation of CpG dinucleotides; hypersensitivity to DNA nucleases and restriction enzymes; and increased activating histone marks, including histone H3/H4 acetylation (Outters et al., 2015). These differences lead the two alleles to be

differentially available for RAG binding and recombination (Vettermann and Schlissel, 2010). The order and significance of these influences is still unclear, and the initial mechanism that dictates the selected allele remains debated. A detailed discussion of competing models can be found here (Vettermann and Schlissel, 2010; Outters et al., 2015).

Positive Selection of B Lymphocytes

The vast pool of antigen receptors allows for modest binding and subsequent detection of most antigens, but once lymphocytes are activated they undergo numerous rounds of selection to increase their affinity for the antigen. We will focus on the positive selection of antigen-selective B lymphocytes in the germinal center of lymph nodes. Of course, stochastic production of TCRs and BCRs can also lead to dangerous autoimmune reactions; these are minimized due to distinct processes of negative selection that occur during lymphocyte development (reviewed in Klein et al., 2014; Nemazee, 2017; Rose, 2017).

Somatic hypermutation (SHM), a key process in affinity maturation, functions to diversify BCRs and promote the adaptive immune response. During SHM, the *BCR* locus undergoes a significant increase in the rate of point mutations compared to the rest of the genome (Forrest and Oprea, 1999). These mutation “hotspots” usually encode the complementarity-determining regions in the variable N-terminus of the antibody that interact with and recognize the antigen. SHM occurs when the enzyme activation induced deaminase (AID) targets mature rearranged V(D)J and switch regions of *Ig* genes (Pilzecker and Jacobs, 2019). AID functions by binding to single-stranded DNA and removing the amino group from cytosine, which produces highly mutagenic deoxy-uracil in the DNA of both *Ig* strands at a high rate. DNA damage response processes then generate base substitutions at and around the lesion created by the deoxy-uracil (Pilzecker and Jacobs, 2019).

Lymph node germinal centers (GCs) are the site of B lymphocyte clonal selection that drives affinity maturation to produce memory B cells and antibody-secreting plasma cells (Victoria and Mesin, 2014). The GC is separated into a dark and light zone. B cells undergo SHM while proliferating in the dark zone (McKean et al., 1984; Victoria and Mesin, 2014). This generates a diverse clonal pool that migrates to the light zone for selection. In the light zone, B cells use their antigen receptors to retrieve antigen from the surface of follicular dendritic cells (FDCs) and then present this antigen to T follicular helper (Tfh) cells to receive survival signals. Tfh cells were found to be the limiting factor in GC selection, as they can only interact with a small portion of the B cells (Victoria et al., 2010). This creates competition between B cell clones to retrieve antigen from FDCs and present it to Tfh cells, with higher affinity BCRs being able to present more antigen and receive the limited Tfh support (Victoria et al., 2010). Tfh cells send support signals in the form of cytokines and cell surface receptors like CD40L, IL-21, and IL-4 to allow B cell survival and migration into the dark zone for further proliferation and SHM (Crotty, 2014). Post-transcriptional regulation of the chemokine CXCL12 receptor CXCR4, along with differential expression of polycomb proteins, have been shown to mediate zonal migration and polarization

(Allen et al., 2004; Okada et al., 2005; Allen et al., 2007a; Allen et al., 2007b; Victora et al., 2010).

Multiple rounds of this selective process produces a robust pool of antibodies that have significantly improved affinity for the antigen. Though selection of B cells in the GC starts out from mostly *interclonal* competition, competition eventually progresses to *intraclonal* competition between variants generated by SHM (Jacob et al., 1993). Some GCs will show clonal dominance of high affinity lineages, but this dominance is not required for high affinity clones to emerge and is not present in all GCs (Tas et al., 2016).

DOWN SYNDROME CELL ADHESION MOLECULES

Neural circuit wiring is an extremely important process that is highly dependent on the proper patterning of neurons within the developing nervous system. While neurons positively select their partners through recognition of deterministically expressed cell surface molecules, neurons also have to avoid synapsing with themselves in order to establish their typical anatomies and heterologous partners. This process, called neuronal self-avoidance, requires neurons to distinguish “self” from “nonself.”

In both vertebrates and insects, neuronal self identity is determined by randomized expression of subsets of possible cell surface molecules. These expression patterns are distinct across individual neurons, even neurons of the same type, and can be thought of as a unique barcode displayed on the surface of each individual cell. In *Drosophila*, the protein family used for this purpose is the Dscam (Down syndrome cell adhesion molecule) family of immunoglobulins. Via alternative splicing, the *Drosophila Dscam1* locus encodes up to 38,016 distinct Dscam isoforms, all of which contain the same basic structure: an ectodomain comprised of 10 immunoglobulin (Ig) domains and six fibronectin type III repeats, a transmembrane domain, and a C-terminal cytoplasmic tail (Figure 4A) (Schmucker et al., 2000). Four variable domains are encoded by clusters of exon variants, which are spliced independently of each other: *Ig2* (12 variants), *Ig3* (48 variants), *Ig7* (33 variants), and the transmembrane domains (2 variants). This means that for 38,016 distinct isoforms there are potentially $(12 \times 48 \times 33 = 19,008)$ distinct ectodomains; at least 18,048 of these ectodomains are confirmed to support “homophilic” binding, or binding between identical isoforms (Wojtowicz et al., 2007).

Homophilic binding of two Dscams generates a repulsive response. When coupled with the immense diversity of Dscam isoforms, which makes it unlikely that neighboring neurons will express identical sets of isoforms, it becomes clear how Dscams mediate neuronal self-avoidance: neurites within the same neurons will express the same Dscams and repel each other, while neurites between neighboring neurons will express different Dscams and allow synapsing. The power of this “barcoding system” is evident from mutation and ablation studies: Where *Dscam1* function is disrupted, dramatic defects in self-recognition are observed, including increases in intraneuronal dendritic

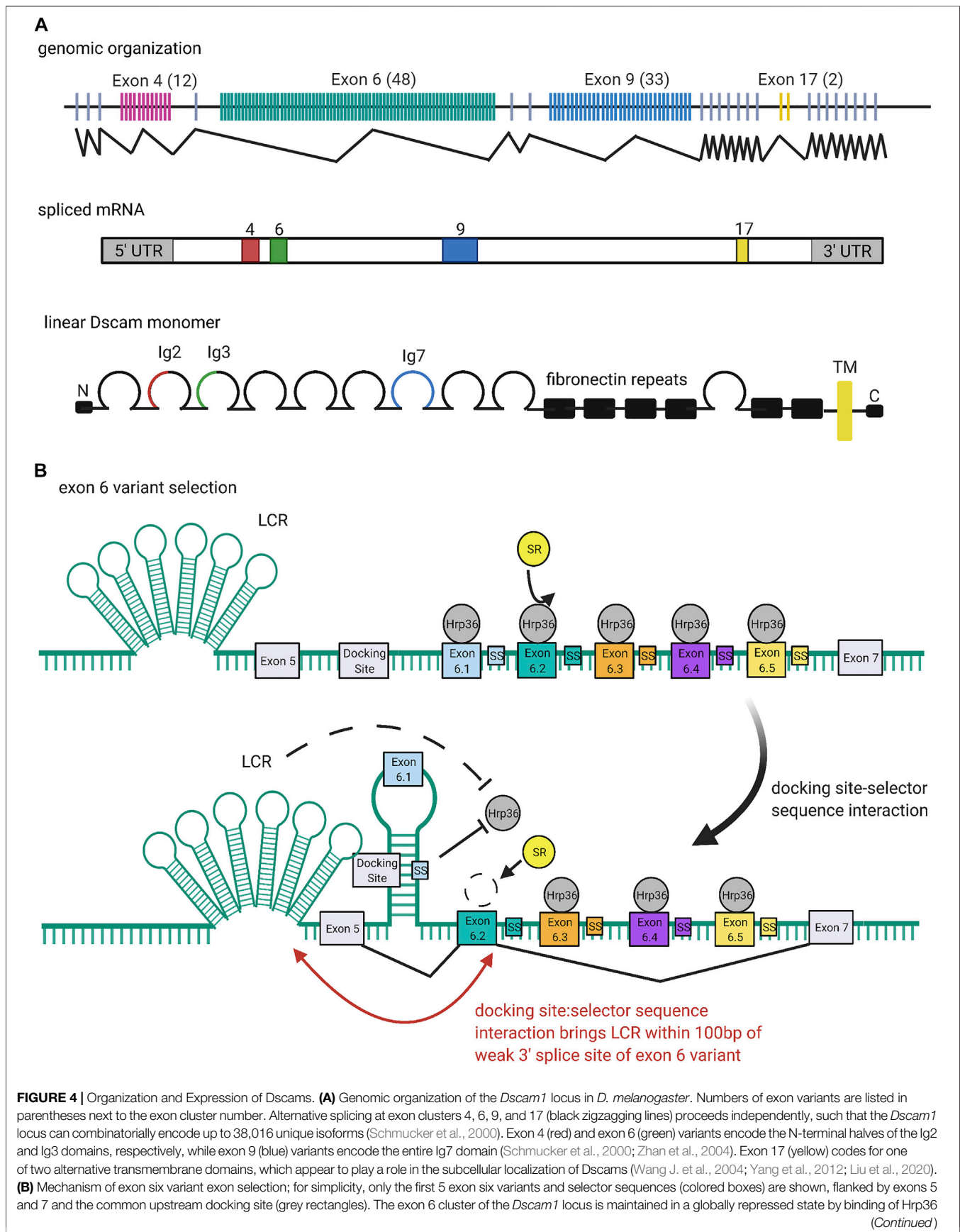


FIGURE 4 | proteins (dark grey circles) to each variant (Olson et al., 2007). Binding of a variant selector sequence to the docking site forms an RNA hairpin loop which prevents inclusion of variants contained within the loop but promotes inclusion of the variant immediately downstream (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011; Hemani and Soller, 2012; Xu et al., 2019). This interaction also brings the locus control region (LCR) within 100 bp of the weak 3' splice site of the downstream variant, which may also promote variant inclusion by facilitating recognition of the splice site (Wang X. et al., 2012). Both the docking-site selector sequence interaction and LCR are thought to promote variant inclusion by antagonizing binding of the repressive Hrp36 proteins, which allows binding of inclusion-promoting SR proteins (yellow circles) at the selected variant (Olson et al., 2007; Wang S.-Z. et al., 2012). Figure inspiration was drawn from various sources (Schmucker et al., 2000; Wang X. et al., 2012).

crossing in dendritic arborization (da) neurons (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007) and failure of sister branch segregation in the axons of mushroom body (MB) neurons (Wang et al., 2002; Hattori et al., 2007). In addition to self-recognition, Dscams have also been suggested to mediate synaptic target selection and axon guidance in several kinds of neurons (Wang et al., 2002; Hummel et al., 2003; Zhan et al., 2004; Zhu et al., 2006; Millard et al., 2010).

Structure and Function of Dscam Homophilic Binding

As discussed, binding specificity is critical to Dscams function in neuronal self-avoidance (Neves et al., 2004; Zipursky et al., 2006). Indeed, both *in vitro* and *in vivo* studies have demonstrated highly specific homophilic binding, to the point that isoforms differing in just a few residues exhibit very weak or no binding (Wojtowicz et al., 2004; Zipursky and Sanes, 2010). How is this exquisite specificity determined? Furthermore, how does attractive homophilic binding generate a repulsive response? Briefly, ELISA binding assays have determined that the 8 N-terminal domains (Ig1-Ig8) of Dscam proteins are sufficient to support normal binding (Figure 5A) (Wojtowicz et al., 2004). Contained in this sequence are the Ig2, Ig3, and Ig7 variant domains, which determine the binding specificity of isoforms by selectively “matching” with their identical counterparts (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Sawaya et al., 2008). While binding of individual variable domains is modular, binding of whole Dscams is all-or-nothing (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Sawaya et al., 2008). That is, the particular identities of the variable domains do not matter as long as they are the *same* between isoforms, as even a minor mismatch between one pair of variable domains is sufficient to totally disrupt binding (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Sawaya et al., 2008).

Upon homophilic ectodomain binding, the C-terminal cytoplasmic tail of Dscam initiates repulsive signaling in the cell, which eventually leads to repulsion between cells expressing identical isoforms (Matthews et al., 2007). Although the mechanism that promotes this repulsion is still poorly understood, studies have identified a few conspicuous binding partners involved in cytoskeletal rearrangement (Hing et al., 1999; Schmucker et al., 2000; Worby et al., 2001). Among these is the adaptor protein Dock, which binds the Dscam cytoplasmic tail and recruits the effector kinase Pak1, which is implicated in several pathways underlying neurite repulsion (Figure 5A) (Hing

et al., 1999). However, while the Dock-Pak axis seems to be necessary for Dscam-mediated axon guidance (Schmucker et al., 2000), loss of Dock or Pak1 does not produce dendrite self-crossing phenotypes (Hughes et al., 2007). This indicates that repulsive signaling required for Dscam-mediated self-avoidance can proceed independently of Dock and Pak.

A recent study investigating DSCAMs, the mammalian homologs of fly Dscams, found that both DSCAMs and Dscams share a predicted nuclear localization signal (NLS) and can be cleaved by proteolysis *in vivo* (Sachse et al., 2019). In mammalian DSCAMs, the NLS was found to promote translocation of cleaved DSCAM cytoplasmic fragments into the nucleus, where they affect expression of genes associated with synapse formation (Figure 5A) (Sachse et al., 2019). Future research should assess whether this signaling mode occurs in *Drosophila*.

Mutually Exclusive Splicing Generates Isoform Diversity

Similar to their cousins in the immunoglobulin superfamily, Dscams rely on large arrays of diverse variants to function (Hattori et al., 2007). But unlike TCRs and BCRs, Dscam isoform diversity is generated at the RNA level. Schmucker and colleagues were the first to note the wide variety of Dscam isoforms in fruit flies (Schmucker et al., 2000). cDNA and genomic analyses of Bolwig's nerves in *D. melanogaster* embryos revealed alternative sequences for the extracellular Ig domains 2, 3, and 7, as well as the transmembrane domain (Schmucker et al., 2000). The N-terminal half of Ig2 is encoded by variants of exon 4, the N-terminal half of Ig3 is encoded by variants of exon 6, the entire Ig7 domain is encoded by variants of exon 9, and the entire transmembrane domain is encoded by variants of exon 17 (Schmucker et al., 2000; Zhan et al., 2004). Exon clusters 4, 6, 9, and 17 were found to have 12, 48, 33, and 2 exon variants, respectively. Each variant is spliced in a mutually exclusive manner, such that each *Dscam* cDNA sequence only contains one of each variable exon 4, 6, 9, and 17 (Figure 4A) (Schmucker et al., 2000). Further, splicing of different exon clusters proceeds independently, which is why the *Dscam1* locus can combinatorially encode up to 38,016 unique isoforms (Schmucker et al., 2000). Combined with the fact that individual neurons express several isoforms simultaneously, *Dscam1* turns out to be a powerful system for uniquely “barcoding” cells (Celotto and Graveley, 2001; Neves et al., 2004; Zhan et al., 2004).

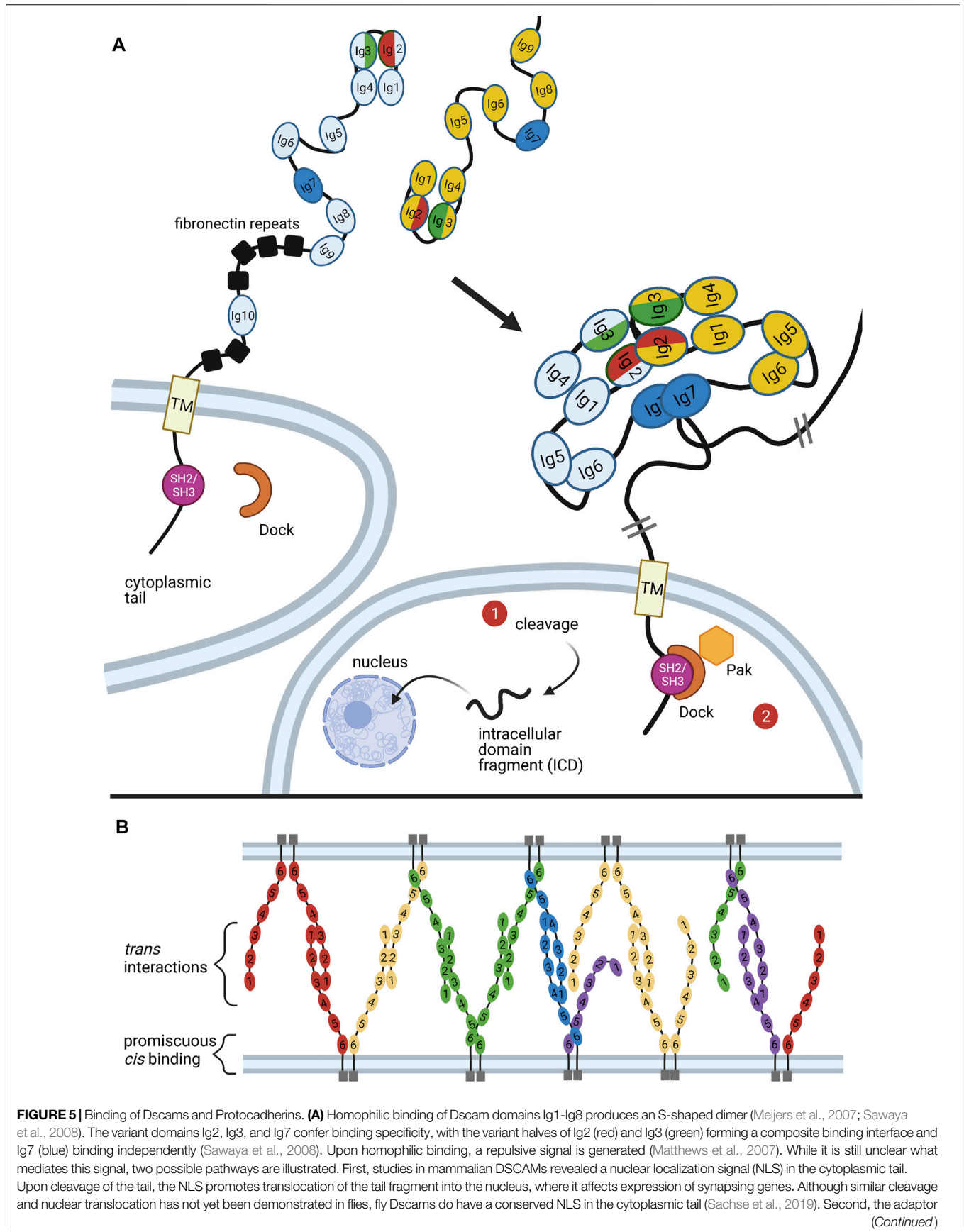


FIGURE 5 | protein Dock has been shown to bind the SH2/SH3 domains of the Dscam cytoplasmic tail and recruit the effector kinase Pak1 (Hing et al., 1999). But while Pak1 has been shown to mediate axon guidance, it does not seem to be necessary for neuronal self-avoidance (Hughes et al., 2007). **(B)** Similar to Dscams, protocadherins bind homophilically. A mixture of *trans* and *cis* interactions forms a zipper-like lattice spanning neighboring cell membranes (Brasch et al., 2019). Figure inspiration was drawn from various sources (Sawaya et al., 2008; Schmucker and Chen, 2009; Guo et al., 2012; Chen and Maniatis, 2013; Goodman et al., 2017).

Variant Exon Inclusion in Alternative Splicing Is Probabilistic

Based on data from *D. melanogaster* exon 4 splicing reporter lines, Miura and colleagues proposed that this mutually exclusive alternative splicing proceeds in a probabilistic fashion (Miura et al., 2013). Within class IV da neurons, they observed that all 12 variants of exon 4 had different yet stable probabilities of inclusion (Miura et al., 2013). In contrast, the probabilities of variant inclusion differed between different classes of neurons. For example, exon 4.2 was expressed in more than half of class IV da neurons, but negligibly expressed in Kenyon cells (Miura et al., 2013). Further, in comparing class IV da neurons in late second and wandering third instar larval stage, Miura and colleagues found that the inclusion frequencies of exon 4 variants changed over time (Miura et al., 2013). Their findings coincide with other studies which demonstrate that *Dscam1* exon selection is biased by developmental stage, tissue type, and even by neuronal subtype (Celotto and Graveley, 2001; Neves et al., 2004; Zhan et al., 2004).

Aside from the general observation that alternative splicing of exon clusters is probabilistic, the mechanism underlying exon variant selection remains to be elucidated; that is, it is still unclear how exon variants are specifically, exclusively, and stochastically selected for inclusion during splicing (Hemani and Soller, 2012). The picture is complicated by the fact that different exon clusters in the *Dscam1* locus seem to employ different methods for mutually exclusive alternative splicing (Hemani and Soller, 2012). To consider one proposed mechanism in detail, this review will focus on mutually exclusive splicing in the exon 6 cluster of the *D. melanogaster Dscam1* locus. More information on regulation of this or other clusters in *Drosophila* or other organisms can be found elsewhere (Graveley, 2005; Anastassiou et al., 2006; Yang et al., 2011; Wang S.-Z. et al., 2012; Hemani and Soller, 2012; Yue et al., 2016a; Yue et al., 2016b; Haussmann et al., 2019; Ustaoglu et al., 2019; Xu et al., 2019).

RNA Secondary Structures Mediate Inclusion of Single Exon Variants

Similar to the selection mechanisms for VSGs and protocadherin exon variants (discussed below), the exon six variants of the *Dscam1* locus are maintained in a repressed state until a selection event specifically activates a variant for expression (**Figure 4B**). In the case of the *Dscam1* locus, the selection event is the formation of RNA secondary structures which antagonize binding of repressive heterogeneous nuclear ribonucleoproteins (hnRNPs) and promote binding of serine-arginine rich (SR) proteins (Olson et al., 2007).

Graveley (2005) first reported conserved sequences within the exon 6 cluster that seem to be required for mutually exclusive selection of exon six variants: a “docking site,” located in an intron upstream of the first exon six variant, and a “selector sequence,” one of which is located directly upstream of each of the

48 exon six variants (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011). The docking site and each selector sequence are predicted to form an RNA stem-loop structure by base-pairing (Graveley, 2005; May et al., 2011). This stem-loop prevents splicing inclusion of the exon variants contained within the loop but promotes specific inclusion of the exon directly downstream of it (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011; Hemani and Soller, 2012; Xu et al., 2019). Because the selector sequences bind to offset, overlapping portions of the docking site, only one selector sequence is predicted to bind, ensuring that there is only one such stem-loop structure at any given time (Graveley, 2005; Anastassiou et al., 2006; May et al., 2011). In addition to competing docking site-selector sequence interactions, it appears that a locus control region (LCR) is necessary for the activation of exon six variants (Wang X. et al., 2012). The LCR is a large tandem stem-loop RNA structure. In *Drosophila* species it forms a highly conserved “hexaleaf” consisting of ~700 bp of scattered upstream intronic sequences (Wang S.-Z. et al., 2012).

RNAi screens by Graveley and colleagues identified Hrp36 (Hrb87F) as the hnRNP responsible for global repression of the exon 6 cluster (Olson et al., 2007). Hrp36 was shown to bind to the exon 6 cluster and is required to repress the inclusion of extra exon six variants. Further, Hrp36 was shown to inhibit binding of SR proteins, which are known to regulate alternative splicing and promote exon inclusion (Olson et al., 2007). Thus, the current model is that an Hrp36 protein binds at each exon six variant within the cluster, maintaining it in a repressed state until an upstream docking site-selector sequence stem-loop somehow dislodges Hrp36 from the proximal variant. The LCR may also help destabilize Hrp36 binding (Wang X. et al., 2012). Dissociation of the Hrp36 protein allows SR proteins to bind the proximal variant and promote its inclusion in splicing (Graveley, 2005; Olson et al., 2007; Hemani and Soller, 2012; Xu et al., 2019). Separately, it has also been suggested that the LCR promotes recognition of weak splice sites in exon variants. Specifically, upon formation of a docking site-selector sequence stem-loop, the LCR is brought within 100 bp of both splice sites of the proximal variant, allowing it to activate inclusion in a proximity-dependent manner (Wang S.-Z. et al., 2012).

To sum, exon inclusion in the *Dscam1* exon 6 cluster appears to be determined by the binding ability of different selector sequences, which may be modulated by splicing factors and RNA-binding proteins (RBPs) such as SR proteins and hnRNPs (or other, noncanonical RBPs, as in the case of the exon 9 cluster, reported elsewhere (Ustaoglu et al., 2019)). It is possible that *deterministic* regulation of these protein factors, which themselves mediate probabilistic events in splicing, underlies the “stochastic yet biased” expression of different exon variants among different cell types and at different times

(Neves et al., 2004; Zhan et al., 2004; Miura et al., 2013). These protein factors may stably associate with splicing machinery in a complex with chromatin, allowing them to sterically exclude exon variants; this may also explain the fact that individual cells only express a finite number of Dscam isoforms (Miura et al., 2013). Future work should investigate the possibility of active negative feedback mechanisms regulating the number of expressed isoforms.

Experimental evidence and comparative genomic analyses indicate that selection of exons 4, 9, and 17 also relies on competing RNA secondary structures, which may be recognized by distinct but overlapping sets of RBPs. This is discussed further elsewhere (Park et al., 2004; Anastassiou et al., 2006; Olson et al., 2007; Yang et al., 2011; Wang X. et al., 2012; Hemani and Soller, 2012; Yue et al., 2016a; Yue et al., 2016b; Haussmann et al., 2019; Ustaoglu et al., 2019; Xu et al., 2019). There is still much to be understood about the mechanisms regulating exon choice within each cluster. It may also prove fruitful to investigate possible crosstalk between the different exon clusters.

Dscam Diversity Is Required for Proper Neural Patterning

Studies that reduced the number of possible Dscams underline the importance of great isoform variety. Regarding self-avoidance, studies found that reducing the Dscam repertoire to just one isoform produced marked neural circuit defects in MB and olfactory receptor (OR) neurons (Hattori et al., 2007; Matthews et al., 2007; Soba et al., 2007). In another study that reduced the isoform repertoire, it was found that flies with at least 4,752 Dscam isoforms were indistinguishable from wild-type controls, while flies with 1,152 isoforms or less demonstrated substantial self-branching defects in *da* neurons (Hattori et al., 2009). These branching defects improved as the number of potential isoforms increased, indicating that self-avoidance requires several thousand different isoforms (Hattori et al., 2009). The finding that neurons require diverse Dscams not only to avoid synapsing with themselves but also to perform anatomic work such as axonal branching suggest that the repulsive force of self-avoidance is used to generate neuronal shapes. How the strength of this force is regulated or differentially harnessed in the production of distinct neuronal shapes is of interest in future work.

While a large variety of isoforms is clearly required, it is unclear whether any one isoform is necessary for normal patterning. In particular, studies that serially deleted different exon 4 variants (thereby eliminating particular Ig2 domains) did not produce observable phenotypes in MB or *da* neurons, indicating that self-avoidance does not require any specific isoform (Wang J. et al., 2004; Hattori et al., 2009). On the other hand, another study reducing diversity to 22,176 isoforms in mechanosensory neurons found defects in axonal branch extension and branching patterns that correlated with particular deletion alleles, suggesting that some connectivity patterns may be mediated by specific isoforms (Chen et al., 2006). It may be that specific isoforms are needed for some types of neural patterning processes, such as axonal targeting and

branching, but not for dendritic self-avoidance. If so, this may also reconcile the bias for certain exon variants at certain developmental stages and within specific cell types: different Dscam isoforms may be required for different developmental and patterning processes (Celotto and Graveley, 2001; Zipursky and Sanes, 2010).

Protocadherins as Analogs for Dscams

Vertebrate protocadherins function analogously to invertebrate Dscams in that both systems mediate neuronal self-avoidance (Garrett et al., 2018). As with invertebrate Dscams, diverse sets of protocadherin isoforms are generated and go on to mediate processes required for proper neural circuit wiring in vertebrates (Schreiner and Weiner, 2010). A notable difference between invertebrate Dscams and vertebrate protocadherins is how variation is produced in each system: alternative splicing generates diverse Dscam isoforms, while utilization of alternative promoters generates diverse protocadherin isoforms (Schmucker et al., 2000; Wojtowicz et al., 2004; Jin et al., 2018). Specifically, Dscams rely on splicing and associated proteins, while PCDHs utilize a CTCF/cohesin-mediated DNA looping mechanism to select proper promoters (Schmucker et al., 2000, p. 200; Wojtowicz et al., 2004). Both mechanisms, however, are prime examples of the ability of non-deterministic events to generate great protein diversity. Such diversity is especially helpful in patterning the nervous system, given that each neuron is likely to have several neighbors, each of which needs to have a different “barcode” if it is to form unique and overlapping networks of connections. Further, the binding of variable domains in either type of protein is exquisitely specific, much like the binding of immunoglobulins and T-cell receptors (**Figure 5B**) (Zipursky and Sanes, 2010). It is remarkable that Dscams and protocadherins have such convergent functions, given the significant differences in their phylogeny, morphology, and mechanistic origins.

While the *DSCAM* gene is conserved in mammals, it does not encode diverse isoforms. In some tissues, including the vertebrate eye, *DSCAM* has been shown to play a deterministic role in synaptic matching (Yamagata and Sanes, 2008). It has also been shown to increase stringency of synaptic partnerships in other areas of the brain (Garrett et al., 2018). However, a recent analysis found that mammalian *DSCAM* and *DSCAML1*, along with other members of the basigin-related family, are sometimes expressed in a mutually exclusive pattern (Iakovlev et al., 2021). It is therefore possible that *DSCAM* genes also contribute to non-deterministic aspects of cell identity in mammals.

PROTODADHERINS

The lack of Dscam diversity in mammals prompted a hunt for protein families performing barcoding or self-avoidance roles in mammals. Work over the last 20 years has shown that a subset of cadherins, the clustered protocadherins, mediates neuronal self-identification in mammals. Just like Dscams, protocadherins generate unique signatures, or “barcodes,” in individual neurons

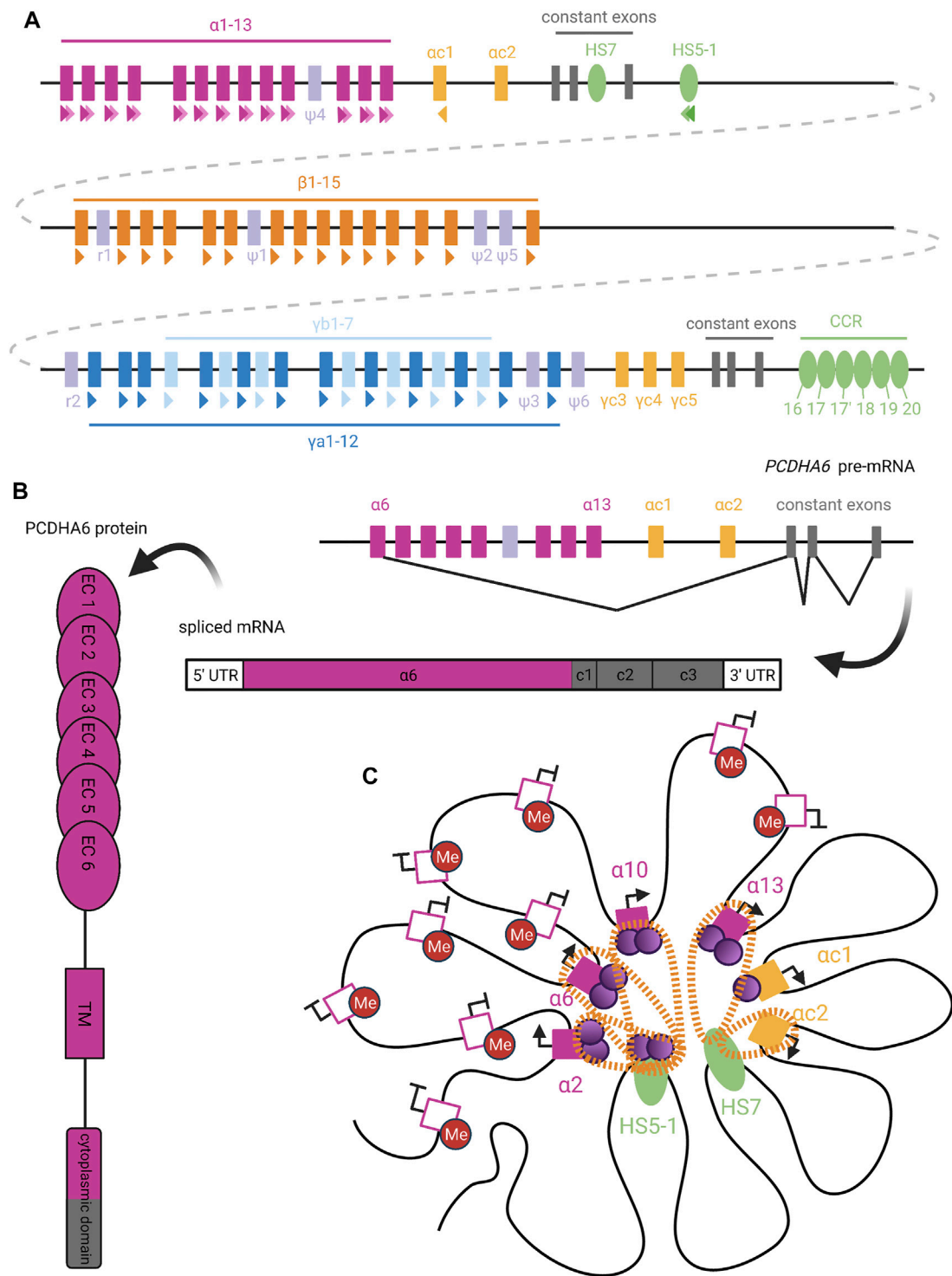


FIGURE 6 | Organization and Expression of Protocadherins. **(A)** Genomic organization of the human clustered protocadherins. In humans, the *PCDHA*, *PCDHB*, and *PCDHG* genes are arranged in tandem clusters, with a few intervening pseudogenes (lavender), in the 5q31 region of chromosome 5 (Wu and Maniatis, 1999). The *PCDHA* and *PCDHG* clusters also have deterministically-expressed terminal V exons (yellow) and constant exons (grey), while the *PCDHB* cluster only contains stochastically-expressed V exons (Tasic et al., 2002). Colored triangles under exons or enhancers denote CSEs, with staggered triangles indicating two CSEs for a given element. **(B)** Upon *PCDHA* promoter selection, transcription proceeds through all downstream V and C exons. Intervening V exons are spliced out such that mature *PCDHA* mRNA contains just one variable exon spliced to three constant exons (Tasic et al., 2002). The variable exon encodes six ectodomains, the transmembrane domain, and the N-terminal half of the cytoplasmic domain in the PCDHA protein; the rest is encoded by the constant exons (Mah and Weiner, 2016). **(C)** (Continued)

FIGURE 6 | Extensive DNA looping between active *PCDHA* promoters (filled pink or yellow squares) and long-range enhancers (HS5-1 and HS7; green ovals) is mediated by complexes of CTCF (purple circles) and cohesin (orange perforated bands) (Kehayova et al., 2011; Guo et al., 2012). In certain cases, binding of cohesin or CTCF alone (as with the *PCDHAC1* and *PCDHAC2* exons, respectively) is able to mediate promoter-enhancer interactions (Guo et al., 2012). To allow CTCF binding, candidate *PCDHA* genes must escape repressive DNA methylation (red circles), perhaps with the help of long non-coding RNA (Canzio et al., 2019). We diagram the potential for co-expression of a few isoforms in the same cell, as has been observed in single-cell analyses (Esumi et al., 2005; Kaneko et al., 2006; Mountoufaris et al., 2017). Figure inspiration was drawn from various sources (Sawaya et al., 2008; Weiner and Jontes, 2013; Massah et al., 2015; Mancini et al., 2020; Wu and Jia, 2021).

and function via homophilic repulsion (Weiner and Jontes, 2013). The regulation of clustered *PCDH* expression shares much in common with other systems discussed in this review: protocadherins utilize enhancers to stochastically select a promoter, similar to olfactory receptors; they contain variable and constant regions analogous to those of immunoglobulins; and their expression is sometimes monoallelic and restricted by heterochromatinization.

Genomic Organization of Clustered Protocadherins

Clustered *PCDH*s are organized into alpha (*PCDHA*), beta (*PCDHB*), and gamma (*PCDHG*) clusters, arranged in tandem along chromosome 5 in humans. The *PCDHA* genes are encoded by a set of 15 large variable (V) exons that precede 3 constant (C) exons (Figure 6A) (Wu and Maniatis, 1999). Each V exon is preceded by its own promoter, and transcription can initiate from any of these 15 promoters. The first transcribed V exon, determined by which promoter is selected, is then spliced to the C exons, removing the intervening V exons. The *PCDHG* cluster is arranged similarly to the *PCDHA* cluster, with 22 V exons preceding 3 C exons (Morishita and Yagi, 2007). The *PCDHB* cluster differs from the other two in that it does not contain any C exons to complement its 22 V exons and is thus a set of distinct genes (Tasic et al., 2002).

The final 2 V exons within the *PCDHA* cluster, *PCDHAC1* and *PCDHAC2*, and the *PCDHGC3*, *PCDHGC4*, and *PCDHGC5* from the *PCDHG* cluster are very similar to each other but not to other V exons (Wu and Maniatis, 1999). These five exons are expressed deterministically and will not be discussed further here.

Stochastic Selection of V Exon Promoters Dictates Isoform Expression

The promoters preceding *PCDH* V exons contain a conserved sequence element (CSE) (Wu et al., 2001), which has been found to serve as an essential promoter binding motif (Tasic et al., 2002). With the exception of *PCDHAC1* and *PCDHAC2*, a second CSE has been observed within the exonic sequence of each *PCDHA* V exon (Figure 6A) (Chen and Maniatis, 2013). Similar CSE regions are also present in transcriptional *cis*-enhancer elements that are specific to each *PCDH* gene cluster (Hirayama and Yagi, 2017): The *PCDHA* cluster enhancer element HS5-1 is found downstream of the cluster's third C exon and contains two CSE sites (Ribich et al., 2006). A long-range *PCDHB* enhancer, known as the clustered control region (CCR), has also been identified downstream of the *PCDHG* cluster (Yokota et al., 2011). Deletion of these enhancers

affected the expression of their respective clusters, with little effect on *PCDHG* cluster expression. This suggests the existence of distinct *PCDHG* enhancer element(s) ($X\gamma$ elements) whose exact genomic location is unknown (Yokota et al., 2011).

In order to initiate transcription at a given V exon, the CSE sites within its promoter region, its respective cluster's enhancer element(s), and exon sequence (for *PCDHA* only), must escape repressive methylation established by DNMT3B during embryogenesis (Garrett et al., 2019). A recent study has implicated antisense long non-coding RNA (as-lncRNA) in the demethylation of promoter and exonic CSE sites within the *PCDHA* cluster (Canzio et al., 2019). This process may facilitate variation in *PCDHA* promoter choice by equalizing interaction of the enhancer with distal and proximal promoters. That is, in the absence of methylation, proximal promoters are preferentially chosen (Canzio et al., 2019). Thus, global methylation of the alpha cluster, alleviated by stochastic demethylation by as-lncRNA, prevents inclusion bias caused by proximity of certain promoters to enhancer elements. The same study did not detect the presence of any as-lncRNA corresponding to the *PCDHB* and *PCDHG* clusters. As such, it remains unclear what mechanisms are responsible for the demethylation of CSE sites within these two clusters.

Once the appropriate CSE sites have been demethylated, the next step of transcription initiation can proceed. The CSEs serve as binding sites for the insulator protein CTCF, which can interact with cohesin to form a CTCF/cohesin complex (Guo et al., 2012). Chromosome conformation capture (3C) assays have demonstrated the ability of CTCF/cohesin complexes to mediate interactions between the V exon promoters and enhancers through DNA-looping (Figure 6C) (Guo et al., 2012). Interestingly, a second *PCDHA* enhancer element (HS7), which is located within the intron between the second and third C exons, lacks a CSE site but is still able to mediate DNA-looping through interaction with cohesin alone (Guo et al., 2012). Nonetheless, several studies have pointed towards CTCF/cohesin complex interactions between V exon promoters and enhancers as a necessary step for expression of most *PCDH* isoforms from all three clusters (Monahan et al., 2012; Hirayama and Yagi, 2017). This conclusion is further supported by the fact that in non-neuronal cell types, competitive binding of the REST/NRSF repressor complex to the HS5-1 enhancer, rather than the CTCF/cohesin complex, led to significantly decreased *PCDHA* expression (Kehayova et al., 2011). Unlike the proposed model for an "enhancer hub" similar to that involved in olfactory receptor choice, it remains unclear how transcription proceeds following the formation of these promoter/enhancer interactions (Guo et al., 2012).

Incredibly, *PCDH* genes are expressed both monoallelically and biallelically. All three *PCDH* clusters show monoallelic, combinatorial expression of the V exons (Esumi et al., 2005; Kaneko et al., 2006; Hirano et al., 2012). However, the 5 C-type variable exons of the alpha and gamma clusters, *PCDHAC1*, *PCDHAC2*, *PCDHGC3*, *PCDHGC4*, and *PCDHGC5*, are all expressed biallelically (Kaneko et al., 2006). Therefore, both the *PCDHA* and *PCDHG* clusters are regulated under different allelic gene regulation mechanisms, which may help to increase neuronal diversity (Kaneko et al., 2006).

Protocadherin expression is not fully monogenic—in at least some neuronal types, expression of >2 non-C-type isoforms together is common (Kehayova et al., 2011). This could result from unstable promoter choice, such that different isoforms are expressed sequentially and still present contemporaneously, or from incomplete dependence on exclusive enhancers. In support of the second model, *PCDHA* promoters display differential dependence on the HS5-1 enhancer; *PCDHA1-5* promoters are only moderately affected by loss of HS5-1 and could be chosen by an additional enhancer (Kehayova et al., 2011). Remarkably, *Dscam* choice in insects is also non-exclusive. Mixing of isoforms, or indeed variation in chosen isoforms over time, should not impede the function of these gene families in self-recognition as long as all parts of the same neuron have the same isoform mix as one another.

Roles in Neural Circuit Development

PCDHG genes are commonly studied and highly expressed in the dendrites of hippocampal neurons where homophilic cell-cell interactions between isoforms facilitate circuit complexity (Molunby et al., 2016). In these neurons, *PCDHG* acts locally to promote arborization via homophilic matching. This was shown by increasing the likelihood of *PCDHG* homophilic interactions using mutations, which subsequently increased dendritic and circuit complexity (Molunby et al., 2016). This is opposite of what would be expected in other cell types, like retinal starburst amacrine cells and Purkinje cells, in which homophilic *PCDH* interactions lead to a clear self-avoidance pattern (Lefebvre et al., 2012). Thus, depending on the cell type, *PCDHG* interactions may lead to attraction, repulsion, or other dendritic arborization signaling events. It has also been shown that the *PCDHA* cluster is involved in these expression patterns and that *PCDHA* and *PCDHG* isoforms work synergistically to facilitate dendritic arborization pattern formation (Molunby et al., 2016; Ing-Esteves et al., 2018). Comparisons of an allelic series mutant support the conclusion that *PCDHA* and *PCDHG* function together in a dose-dependent and cell-type-specific manner to provide a critical threshold for *PCDH* activity. Although this does create a type of redundancy in *PCDH* stochastic expression, having both *PCDHA* and *PCDHG* expressed is critical for neuronal development (Ing-Esteves et al., 2018). *PCDHB* isoforms, similar to *PCDHA* and *PCDHG*, form *trans*-homophilic interactions (interactions with identical molecules on other cells), but expression patterns are not well classified (Yokota et al., 2011; Hasegawa et al., 2016).

Structural Characterizations

Like *Dscams*, clustered protocadherins bind homophilically, and all the isoforms expressed in a cell must find a match in *trans* in

order to induce strong enough binding to initiate downstream signaling cascades and cellular responses. Recent structural characterizations of the gene family show that the structure of the *PCDH* ectodomain is a zipper-like lattice formed by alternating *cis*- and *trans*-interactions (Figure 5B). As the protein extends out of the cell and interacts with the proteins of adjacent cells, a larger two-dimensional structure is created between the cell membranes (Brasch et al., 2019). This structure is the basis for the initial self-recognition step in neuronal self-avoidance (Brasch et al., 2019). Once these structures are formed, members of all three *PCDH* clusters can mediate highly specific homophilic recognition, maximizing the most favorable protein interactions. This favorable homophilic interaction is maximized when identical isoforms of clustered *PCDHs* are present. For example, neurons expressing five isoforms prefer to form homophilic aggregates with neurons expressing a full identical set of five isoforms rather than those expressing just three or four out of the five. Thus, self-recognition between different neurons is avoided by the expression of a single mismatched isoform (Thu et al., 2014). If there is a perfect match between isoforms (i.e. branches of the same Soma), repulsion of the neurite branches will occur to avoid overlapping. However, the downstream signaling mechanisms mediating repulsion are not yet understood.

OLFACTORY RECEPTORS

Similar to the protocadherin family, mammalian olfactory receptor genes are also expressed monoallelically and monogenically in neurons. Olfactory receptor (OR) genes are organized in large clusters in the genome and, like *VSG* genes or the *PCDHB* cluster, are each independent transcriptional units, not sets of overlapping possibilities like the *BCR/TCR*, *Dscam1*, and *PCDHA* and *PCDHG* loci. ORs are transmembrane chemoreceptors, found on the cell membranes of olfactory sensory neurons (OSNs), which recognize odorant molecules to detect smells. Buck and Axel correctly estimated that there are roughly 1,000 OR genes in the mammalian genome (Buck and Axel, 1991). Stochastic expression of individual OR genes across individual olfactory sensory neurons allows each neuron to serve as a sensor for a limited repertoire of odors, simplifying higher-order interpretation of olfactory information.

The mammalian olfactory system is composed of the main and accessory olfactory systems (Spehr and Munger, 2009). The main olfactory epithelium (MOE), located in the main olfactory system in the nasal cavity, contains the OSNs, which synapse onto the main olfactory bulb. The accessory olfactory system, which is located in the vomeronasal organ in the nasal cavity, contains OSNs that express vomeronasal receptors and is absent in humans and other primates. A further look into the ORs of mice and humans reveals that they share many subfamilies of OR genes. However, homology relationships are difficult to discern at the gene level, and mouse subfamilies typically include more OR genes than human subfamilies (Godfrey et al., 2004). This variation in the olfactory receptor repertoire occurs through a

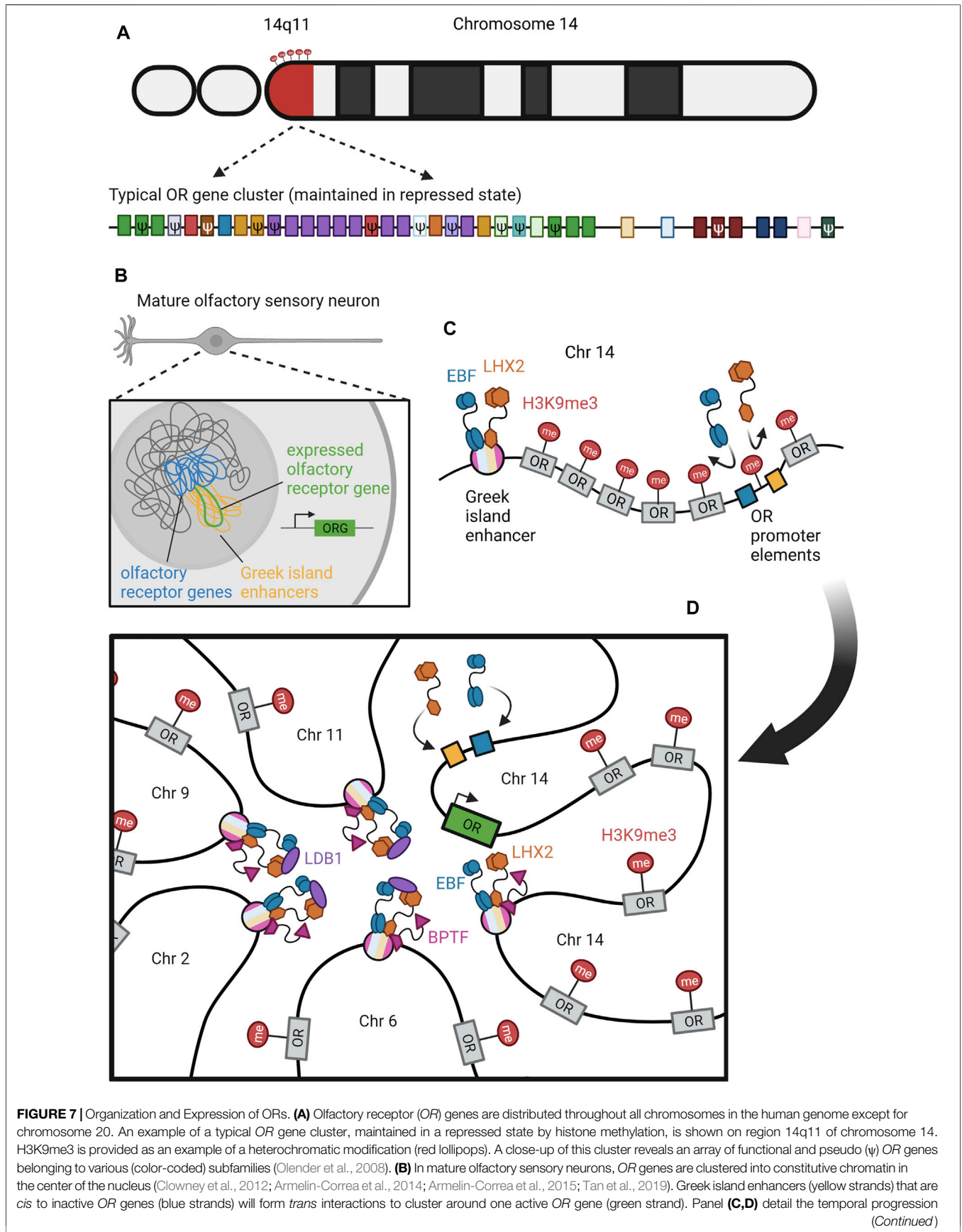


FIGURE 7 | Organization and Expression of ORs. (A) Olfactory receptor (*OR*) genes are distributed throughout all chromosomes in the human genome except for chromosome 20. An example of a typical *OR* gene cluster, maintained in a repressed state by histone methylation, is shown on region 14q11 of chromosome 14. H3K9me3 is provided as an example of a heterochromatic modification (red lollipops). A close-up of this cluster reveals an array of functional and pseudo (ψ) *OR* genes belonging to various (color-coded) subfamilies (Olender et al., 2008). **(B)** In mature olfactory sensory neurons, *OR* genes are clustered into constitutive chromatin in the center of the nucleus (Clowney et al., 2012; Armelin-Correa et al., 2014; Armelin-Correa et al., 2015; Tan et al., 2019). Greek island enhancers (yellow strands) that are *cis* to inactive *OR* genes (blue strands) will form *trans* interactions to cluster around one active *OR* gene (green strand). Panel **(C,D)** detail the temporal progression (Continued)

FIGURE 7 | leading to this arrangement. **(C)** Although Greek island enhancers can make *cis* OR genes competent for activation, they cannot completely relieve heterochromatic repression (red lollipops) (Serizawa et al., 2003; Nishizumi et al., 2007; Khan et al., 2011). Transcription factors BPTF (pink), EBF (blue), and LHX2 (orange) can bind to juxtaposed motifs within enhancers (striped circles), but are not strong enough to overcome heterochromatic repression of the same motifs in OR promoters (yellow and blue boxes) (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2017). **(D)** *Trans* interactions between Greek island enhancers on different chromosomes select one competent OR gene for activation (green box) (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2017; Monahan et al., 2019; Tan et al., 2019). Long-range, interchromosomal interactions are mediated by LDB1 (purple) in concert with EBF and LHX2 (Monahan et al., 2019). These interactions relieve repression on the promoter *cis* to the selected OR, allowing EBF and LHX2 to bind and promote transcription (Monahan et al., 2017). Figure inspiration was drawn from various sources (Malnic et al., 2004; Olender et al., 2008; Monahan et al., 2017; Bashkirova and Lomvardas, 2019; Monahan et al., 2019; Spitz, 2019) (Malnic et al., 2004; Olender et al., 2008; Monahan et al., 2017; Bashkirova and Lomvardas, 2019; Monahan et al., 2019; Spitz, 2019).

process of birth-and-death evolution, and species-level differences are thought to arise through a combination of adaptation to different niches and genetic drift (Nozawa and Nei, 2009). Selection on any one member of such a large gene family is expected to be weak.

Like *Dscam* isoforms, ORs are expressed through a biased stochastic process: While expression of each OR is restricted to a particular zone of the olfactory epithelium, within the zone each olfactory sensory neuron randomly expresses just one allele of one of the ~1000 OR genes. Individual olfactory sensory neurons don't "know" in advance which receptor they will express. This stochastic expression system may facilitate evolutionary turnover at the receptor level as newly born ORs can be expressed privately in existing cell types.

The axons of OSNs expressing the same receptor converge onto glomeruli in the olfactory bulb, which act as a functional unit of odor coding (Chesler et al., 2007). Since a given odorant activates multiple OR species and a given OR responds to multiple odorants, this results in received odorants producing unique combinations of activated glomeruli with varying magnitudes of activity in the olfactory bulb (Malnic et al., 1999). This ensures precise odorant perception and allows the mouse olfactory system, for example, to detect orders-of-magnitude more odors than the number of receptors encoded in the genome (Brochtrup and Hummel, 2011).

Monoallelic and Monogenic Expression of ORs

The expression of just a single olfactory receptor gene per OSN plays two critical roles in odor perception: First, it ensures that each OSN senses just a small set of odorant molecules; second, it allows olfactory sensory neurons to target specific glomeruli in the olfactory bulb. Multiple studies of differentiated OSNs have demonstrated that OR genomic regions remain intact throughout development (Eggan et al., 2004; Li et al., 2004). This is unlike the genetic recombination mechanisms we see in VSG and immunoglobulin diversity, which rearrange genomic loci. Further sequencing data confirmed that ORs are each encoded by individual genes rather than variable exons like *Dscam* and *Pcdh* genes (Li et al., 2004). Not only that, but a single OR gene is selected via the interaction of multiple enhancers (Figure 7), followed by feedback inhibition of remaining ORs (Monahan and Lomvardas, 2015). Investigations into OR selection have therefore focused on epigenetic and transcriptional regulators in an attempt to understand this stochastic system (Shykind, 2005).

In addition to being expressed monogenically (i.e. one OR per neuron), ORs are also expressed monoallelically. It is not clear if this is functionally important or if it is a side effect of the monogenic expression system. If the neuron treats all OR genes as identical and makes a single choice, then this would result in monoallelic as well as monogenic expression. However, previous studies have suggested that monoallelic silencing is a separate process from monogenic OR choice, and that it occurs earlier in development via DNA methylation (Chess et al., 1994; Armelin-Correa et al., 2014).

Transcription of multiple OR genes per OSN precedes the eventual translation of just a single OR per neuron (Hanchate et al., 2015; Tan et al., 2015). Immature neurons may express multiple receptors at low concentrations in overlapping regions of the nasal epithelium (Hanchate et al., 2015). This is reminiscent of the VSG system, in which cells may transcribe multiple VSGs before choice stabilization, and may implicate RNA-based mechanisms of gene silencing and/or feedback mechanisms informing the cell that successful translation has occurred.

Heterochromatic Silencing of Non-chosen ORs

Despite coexpression of multiple OR genes in some immature OSNs, only one functional OR protein is translated in mature OSNs, and the rest of the gene family is transcriptionally silenced. Similar to many other surface molecules previously described, epigenetic regulation is essential for OR choice. Formation of heterochromatin throughout OR gene clusters likely occurs in an early stage preceding OR gene expression, although the extent of heterochromatic silencing in an OSN's ground state is still undetermined (Magklara et al., 2011; Nagai et al., 2016). OR genes are marked by the constitutive heterochromatin marks H3K9me3 and H4K20me3, which is a common characteristic amongst other monoallelic expression patterns (Figures 7C,D) (Magklara et al., 2011). Histone methyltransferases G9a and GLP assist in H3K9 trimethylation and have been shown to be essential for promoting OR diversity, but their significance in the timeline of OSN development has not been explicitly explained (Ferreira et al., 2014; Lyons et al., 2014). Since OR genes are widely repressed by trimethylation, histone demethylases are required to remove constitutive repressive marks from heterochromatin in order to permit expression of the chosen OR (Magklara et al., 2011). The enzymes that demethylate each heterochromatin mark to its dimethylated state have yet to be discovered, but one H3K9me2 demethylase, LSD1, is required for OR transcription (Lyons et al., 2013).

Heterochromatinization of olfactory receptor loci likely allows their reorganization into dense nuclear aggregates and therefore prevents most alleles from contacting the transcriptional machinery (Magklara et al., 2011; Clowney et al., 2012; Le Gros et al., 2016). While heterochromatin is found at the periphery of most mammalian cell nuclei, OSNs and some other types of neurons display a peculiar architecture with constitutive heterochromatin concentrated at the center of the nucleus (**Figure 7B**) (Solovei et al., 2009; Clowney et al., 2012; Armelin-Correa et al., 2015; Tan et al., 2015). Moreover, *OR* genes from all of the chromosomal arrays coalesce into interchromosomal aggregates during OSN development (Clowney et al., 2012; Monahan et al., 2019; Tan et al., 2019). Dismantling these aggregates allows expression of multiple *OR*s per OSN, even when heterochromatin marks remain intact (Clowney et al., 2012). These aggregates are also known to be disrupted in COVID-induced anosmia (Zazhytska et al., 2021).

Enhancer Activation

Reminiscent of *PCDH* promoter selection, enhancer activation is essential for *OR* gene expression. There is a large family of enhancers, called Greek islands, each serving a necessary function in activating their cluster of *cis* *OR*s (Serizawa et al., 2003; Nishizumi et al., 2007; Khan et al., 2011; Markenscoff-Papadimitriou et al., 2014). Additionally, these enhancers interact with one another in *trans* to form a large enhancer cluster that associates with a single *OR*, selecting it for activation (**Figure 7D**) (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2017; Monahan et al., 2019; Tan et al., 2019). These enhancer-*OR* clusters are specific to OSNs, and are not found in other olfactory epithelial cells (Tan et al., 2019). The enhancer regions contain homeodomain binding motifs that are essential for their regulatory function (Nishizumi et al., 2007; Monahan et al., 2019). In transgenic mice, the accumulation of these homeodomain binding motifs upstream of an *OR* promoter led to an increasing probability of reporter transcription (Vassalli et al., 2011). These observations suggest that the accumulation of multiple enhancers provides this homeodomain enrichment to promote *OR* transcription.

The dynamics of these clusters follow a pattern in which the *cis* interactions form first, and then aggregate with other *cis* clusters on separate chromosomes to form larger *trans* complexes (**Figures 7C–D**) (Monahan et al., 2019). The formation of the *OR* foci also plays an important role in bringing *trans* enhancers and *OR* genes together. As such, inhibition of *OR* foci leads to a decrease in enhancer cluster formation and association with *OR* genes, causing a decrease in *OR* expression (Clowney et al., 2012; Markenscoff-Papadimitriou et al., 2014). Recent single cell Hi-C analysis has revealed that the number of enhancers needed to activate a given *OR* is rather low compared to the total number available, with active *OR* genes being associated with an average of six enhancers (Tan et al., 2019).

Several integral proteins have been characterized as key facilitators of cluster formation and *OR* activation. BPTF and LDB1 bind the enhancers and facilitate the aggregation of large enhancer-*OR* complexes to promote *OR* expression

(Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2019). Additionally, LHX2 binds homeodomains (Hirota and Mombaerts, 2004; Rothman et al., 2005; Monahan et al., 2017; Monahan et al., 2019) and the Olf/EBF family bind O/E binding sites (**Figures 7C–D**) (Wang et al., 1997; Wang S. S. et al., 2004), both of which are found on the *OR* promoters and enhancers. The deletion of these transcription factors reduces *OR* expression, and the Olf/EBF family is tightly regulated by Ffp433/OAZ throughout OSN development (Cheng and Reed, 2007; Roby et al., 2012).

Taken together, these discoveries support a model in which each *OR* enhancer first associates in *cis* with a single *OR* gene to promote transcriptional competence, after which the nuclear localization of *OR* foci facilitates the aggregation of these *cis* enhancer-*OR* pairs into larger enhancer complexes (Bashkirova and Lomvardas, 2019). Stochastic association of this large enhancer complex with one of the *OR* genes selected by its *cis* enhancer then promotes its transcription via the accumulated recruitment of transcription factors. Several of these activating clusters emerge in a given OSN nucleus, necessitating a feedback mechanism to silence non-selected *OR* genes and stabilize the selected gene (Tan et al., 2019).

Negative Feedback Regulation

In order to achieve proper OSN identity, *OR* gene silencing is necessary to ensure only one *OR* will function once the neuron has fully matured (Lewcock and Reed, 2004; Fleischmann et al., 2008; Zhou and Belluscio, 2012). Once a functional *OR* is translated and expressed at robust levels, a negative feedback mechanism called the unfolded protein response (UPR) is triggered (Dalton et al., 2013). During this response, the successful and abundant translation of an *OR* leads to downstream signaling that turns off important translation-initiation factors, ceasing translation of most other *OR*s while also inducing expression of adenylyl cyclase type 3 (ADCY3), a molecule that is essential for odor signaling (Ron and Walter, 2007; Wang S.-Z. et al., 2012; Dalton et al., 2013). This feedback mechanism is very similar to the process that takes place within B cells that confirms successful translation of a functional BCR (Hetz et al., 2020). Many chaperones are also essential for *OR* singularity by coordinating the UPR between organelles and the cell membrane (Ron and Walter, 2007; Sharma et al., 2017). Evidence from transgene studies suggests there are also *cis*-regulatory elements in the *OR* coding region that contribute to monogenic expression, but these elements have yet to be identified (Nguyen et al., 2007).

G protein signaling also plays a role in stabilizing *OR* choice following expression of the *OR* protein in the cell membrane. However, it is undetermined whether the chosen *OR* is the first to be expressed or the most robustly expressed. Nevertheless, the selected *OR* protein acts as a GPCR to activate G_{olf} to signal its presence as a functional *OR*. Specifically, the beta-gamma subunit of G_{olf} has been shown to contribute to *OR* gene silencing by heterochromatin regulation in zebrafish, while other model systems have pointed to the importance of the alpha subunit of G_{olf} and its mediation by RIC8B (Von Dannecker et al., 2006;

Ferreira et al., 2014; Machado et al., 2017). Regardless of each subunit's role, G_{olf} then activates ADCY3, which in turn downregulates the aforementioned histone demethylase LSD to prevent activation of other *OR* genes and stabilize *OR* choice (Imai et al., 2006; Dalton et al., 2013; Lyons et al., 2013). The opposing roles of ADCY3 and LSD1 are further reviewed in (Monahan and Lomvardas, 2015).

Altogether, there are multiple known mechanisms of *OR* gene expression and silencing that are essential for monoallelic and monogenic expression. Achievement of singularity has been shown to involve heterochromatin formation, nuclear compartmentalization, enhancer interactions, and a feedback inhibition pathway. However, the temporal and molecular overlaps between these pathways are not yet understood. More research is needed to outline the developmental timeline of *OR* expression and identify causal relationships between these mechanisms that ensure *OR* singularity. This diversity and singularity of *OR* expression facilitates the effective axon guidance and complex circuitry that allows us to detect such a wide spectrum of odors.

SIMILARITIES AND DIFFERENCES ACROSS SYSTEMS

The major shared feature of the five systems covered in this review is their selection of a specific genetic option among a collection of many alternatives in order to express a surface protein that is integral to the cell's identity. While each system has evolved unique mechanisms of monogenic and/or monoallelic expression, there are many parallels in the strategies adopted to accomplish selection, expression, and stabilization of the selected sequence (Table 1).

First, generation of combinatorial products is a key feature in many of these systems that expands the number of potential surface molecules relative to the number of coding segments encoded in the genome. The antigen receptors utilize V(D)J recombination, clustered *PCDH* genes select variable segments, and *Dscam1* splicing combines variable exons to create a massive repertoire of possible combinations from a small number of gene segments. *VSG* genes similarly combine pre-existing *VSG* gene segments to create new mosaic *VSGs*, and, unlike the other systems, these repertoires are constantly evolving and growing as new *VSGs* are made and stored.

These systems also exhibit context-driven restrictions and biases regarding which genes or gene segments are chosen. Within the olfactory epithelium, OSNs are spatially segregated into different zones, and each *OR* is expressed only within a single zone. *Dscams* and *PCDHs* show a similar restriction, but based on cell type rather than spatial location, as the pattern of variable domains expressed appears to vary between different populations of neurons. *VSG* expression instead follows a temporal bias for which the probability of certain *VSGs* appears to change throughout the course of infection, with mosaic *VSGs* emerging in the later stages of infection. Within the confines of these spatial, cell type, and temporal restrictions the available genes and gene segments

still seem to be stochastically distributed. The mechanisms that restrict stochasticity are largely unknown.

The process determining which *VSG* expression site is active relies heavily on nuclear localization and chromatin remodeling, a feature shared by the other four systems. During *VSG*, *OR*, and antigen receptor selection the inactive alleles are repressed in heterochromatic regions, with *VSG* and antigen receptor loci localized to the nuclear periphery and *OR* gene arrays compacted into repressive foci. The selected allele in these systems is found in accessible euchromatic regions, which for antigen receptors is in the nuclear center, for *ORs* entails the active gene escaping the repressive foci, and for *VSG* genes requires the single active expression site to localize to the expression body. These nuclear localization events are intimately tied to chromatin modifications and DNA methylation. *VSG* expression sites localized to the expression body show significant nucleosome depletion. Interestingly, the active *VSG* expression site is enriched with H3K10 acetylation as well as H3K4 trimethylation. H3K4me3 is an activating signature that is also found in the active regions of the *OR*, *PCDH*, and antigen receptor loci. Histone modifications are prominent features of these systems. The antigen receptor loci are further activated by H3K9 acetylation and H3K36 trimethylation, whereas the inactive locus is enriched for repressive H3K9 di- and trimethylation. H3K9 methylation is also utilized to repress *OR* loci alongside H4K20 di- and trimethylation.

Singular *cis*-acting elements are also a common feature activating the selected sequences throughout many of these systems. Antigen receptors and *PCDH* loci share mechanistically similar downstream enhancer elements that loop back to the loci and activate the selected promoter, in both cases scanning for the first available promoter sequence either at the most proximal V segment or hypomethylated *PCDH* segment. Enhancers play a role not only in inducing expression of the selected *OR* gene, but also in carrying out the selection process, as enhancers coalesce first in *cis*, and then in *trans* onto a given *OR* gene for selective expression. Even though *Dscam* selection occurs at the RNA level, an enhancer-like element (the docking site-selector sequence stem-loop) occurs in the secondary structure of the transcript. Reminiscent of the previous systems, the stochastic binding of variant selector sequences to the docking site (which forms the stem-loop) dictates the pattern of exon splicing. While *VSG* expression has not been shown to require enhancer interactions, the selected *VSG* still must be located within the active expression site. Further research into whether a limiting enhancer selects a *VSG* for selection at a particular time will be of great interest.

CONCLUSION

In this review, we have emphasized the importance of stochastic processes in achieving variability in barrier and nervous systems. By analyzing immune system evasion by the parasite *Trypanosoma brucei*, immune system pathogen identification by B cell and T cell receptors, neuronal self-avoidance through diverse expression and self-binding processes of *Dscams* and protocadherins, and the perception of stimuli through the olfactory system, we have shown that the role of diversity is crucial to the survival and function of organisms through both internal and environmental

interactions. Common processes such as monoallelic expression, epigenetic regulation, and specificity of binding of variable domains were discussed. It is important to recognize that all five systems studied are examples of surface molecules that utilize the stochastic selection of genetic material to create cellular diversity. Like aleatoric music and the world-generating algorithms of sandbox games, these systems maximize limited genetic information to construct complex and unpredictable outcomes.

AUTHOR CONTRIBUTIONS

Writing—Original Draft Preparation: all authors.
Writing—Review & Editing: DW, VS, MH, CB, and EJC.

REFERENCES

- Allen, C. D. C., Ansel, K. M., Low, C., Lesley, R., Tamamura, H., Fujii, N., et al. (2004). Germinal center Dark and Light Zone Organization Is Mediated by CXCR4 and CXCR5. *Nat. Immunol.* 5, 943–952. doi:10.1038/ni1100
- Allen, C. D. C., Okada, T., and Cyster, J. G. (2007a). Germinal-center Organization and Cellular Dynamics. *Immunity* 27, 190–202. doi:10.1016/j.immuni.2007.07.009
- Allen, C. D. C., Okada, T., Tang, H. L., and Cyster, J. G. (2007b). Imaging of Germinal center Selection Events during Affinity Maturation. *Science* 315, 528–531. doi:10.1126/science.1136736
- Allshire, R. C., and Madhani, H. D. (2018). Ten Principles of Heterochromatin Formation and Function. *Nat. Rev. Mol. Cell Biol.* 19, 229–244. doi:10.1038/nrm.2017.119
- Alsford, S., and Horn, D. (2012). Cell-cycle-regulated Control of VSG Expression Site Silencing by Histones and Histone Chaperones ASF1A and CAF-1b in *Trypanosoma Brucei*. *Nucleic Acids Res.* 40, 10150–10160. doi:10.1093/nar/gks813
- Alsford, S., Kawahara, T., Isamah, C., and Horn, D. (2007). A Sirtuin in the African Trypanosome Is Involved in Both DNA Repair and Telomeric Gene Silencing but Is Not Required for Antigenic Variation. *Mol. Microbiol.* 63, 724–736. doi:10.1111/j.1365-2958.2006.05553.x
- Anastassiou, D., Liu, H., and Varadan, V. (2006). Variable Window Binding for Mutually Exclusive Alternative Splicing. *Genome Biol.* 7, R2. doi:10.1186/gb-2006-7-1-r2
- Aresta-Branco, F., Erben, E., Papavasiliou, F. N., and Stebbins, C. E. (2019a). Mechanistic Similarities between Antigenic Variation and Antibody Diversification during *Trypanosoma Brucei* Infection. *Trends Parasitology* 35 (4), 302–315. doi:10.1016/j.pt.2019.01.011
- Aresta-Branco, F., Pimenta, S., and Figueiredo, L. M. (2016). A Transcription-independent Epigenetic Mechanism Is Associated with Antigenic Switching in *Trypanosoma Brucei*. *Nucleic Acids Res.* 44 (7), 3131–3146. doi:10.1093/nar/gkv1459
- Aresta-Branco, F., Sanches-Vaz, M., Bento, F., Rodrigues, J. A., and Figueiredo, L. M. (2019b). African Trypanosomes Expressing Multiple VSGs Are Rapidly Eliminated by the Host Immune System. *Proc. Natl. Acad. Sci. USA* 116 (41), 20725–20735. doi:10.1073/pnas.1905120116
- Armelin-Correa, L. M., Gutiyama, L. M., Brandt, D. Y. C., and Malnic, B. (2014). Nuclear Compartmentalization of Odorant Receptor Genes. *Proc. Natl. Acad. Sci.* 111, 2782–2787. doi:10.1073/pnas.1317036111
- Armelin-Correa, L. M., Nagai, M. H., Silva, A. G. L., and Malnic, B. (2015). Nuclear Architecture and Gene Silencing in Olfactory Sensory Neurons. *Bioarchitecture* 4, 160–163. doi:10.4161/19490992.2014.982934
- Backhaus, O. (2018). “Generation of Antibody Diversity,” in *Book: Antibody Engineering*. doi:10.5772/intechopen.72818
- Preparation of figures: VS. Funding acquisition: EJC. DW, VS, and MH share first authorship. This review was written in the Fall 2020 semester by students enrolled in a course at the University of Michigan, MCDB464: Cellular Diversity in the Immune and Nervous Systems.

FUNDING

The preparation of this review was supported by the University of Michigan Program in Neuroscience, by the Rita Allen Foundation Milton Cassell Scholar Award (to EJC), and by NIDCD (Grant number 1R01DC018032, to EJC).

- Barbet, A. F., and Kamper, S. M. (1993). The Importance of Mosaic Genes to Trypanosome Survival. *Parasitol. Today* 9, 63–66. doi:10.1016/0169-4758(93)90039-I
- Bashkirova, E., and Lomvardas, S. (2019). Olfactory Receptor Genes Make the Case for Inter-chromosomal Interactions. *Curr. Opin. Genet. Development* 55, 106–113. doi:10.1016/j.gde.2019.07.004
- Bébin, A.-G., Carrion, C., Marquet, M., Cogné, N., Lecardeur, S., Cogné, M., et al. (20101950). *In Vivo* Redundant Function of the 3' IgH Regulatory Element HS3b in the Mouse. *J. Immunol.* 184, 3710–3717. doi:10.4049/jimmunol.0901978
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., et al. (2005). The Genome of the African Trypanosome *Trypanosoma Brucei*. *Science* 309, 416–422. doi:10.1126/science.1112642
- Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M., and Cooper, M. D. (2012). VLR-based Adaptive Immunity. *Annu. Rev. Immunol.* 30, 203–220. doi:10.1146/annurev-immunol-020711-075038
- Boothroyd, C. E., Dreesen, O., Leonova, T., Ly, K. I., Figueiredo, L. M., Cross, G. A. M., et al. (2009). A Yeast-Endonuclease-Generated DNA Break Induces Antigenic Switching in *Trypanosoma Brucei*. *Nature* 459, 278–281. doi:10.1038/nature07982
- Brandenburg, J., Schimanski, B., Nogoceke, E., Nguyen, T. N., Padovan, J. C., Chait, B. T., et al. (2007). Multifunctional Class I Transcription in *Trypanosoma Brucei* Depends on a Novel Protein Complex. *EMBO J.* 26, 4856–4866. doi:10.1038/sj.emboj.7601905
- Brasch, J., Goodman, K. M., Noble, A. J., Rapp, M., Mannepalli, S., Bahna, F., et al. (2019). Visualization of Clustered Protocadherin Neuronal Self-Recognition Complexes. *Nature* 569, 280–283. doi:10.1038/s41586-019-1089-3
- Brochtrup, A., and Hummel, T. (2011). Olfactory Map Formation in the *Drosophila* Brain: Genetic Specificity and Neuronal Variability. *Curr. Opin. Neurobiol.* 21, 85–92. doi:10.1016/j.conb.2010.11.001
- Buck, L., and Axel, R. (1991). A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition. *Cell* 65, 175–187. doi:10.1016/0092-8674(91)90418-X
- Bühler, M., Paillusson, A., and Mühlemann, O. (2004). Efficient Downregulation of Immunoglobulin mRNA with Premature Translation-Termination Codons Requires the 5'-half of the VDJ Exon. *Nucleic Acids Res.* 32, 3304–3315. doi:10.1093/nar/gkh651
- Canzio, D., Nwakeze, C. L., Horta, A., Rajkumar, S. M., Coffey, E. L., Duffy, E. E., et al. (2019). Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin α Promoter Choice. *Cell* 177, 639–653. e15. doi:10.1016/j.cell.2019.03.008
- Celotto, A. M., and Graveley, B. R. (2001). Alternative Splicing of the *Drosophila* Dscam Pre-mRNA Is Both Temporally and Spatially Regulated. *Genetics* 159, 599–608. doi:10.1093/genetics/159.2.599
- Cestari, I., and Stuart, K. (2015). Inositol Phosphate Pathway Controls Transcription of Telomeric Expression Sites in *Trypanosoma*. *Proc. Natl. Acad. Sci. USA* 112, E2803–E2812. doi:10.1073/pnas.1501206112

- Cestari, I., and Stuart, K. (2018). Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes. *Cg* 19, 119–132. doi:10.2174/1389202918666170911161831
- Chang, C. X. L., Dai, L., Tan, Z. W., Choo, J. A. L., Bertoletti, A., and Grotenbreg, G. M. (20113014). Sources of Diversity in T Cell Epitope Discovery. *Front. Biosci.* 16, 3014–3035. doi:10.2741/3895
- Charles A Janeway, J., Travers, P., Walport, M., and Shlomchik, M. J. (2001). *The Generation of Diversity in Immunoglobulins*. 5th Ed. New York: Immunobiol. Immune Syst. Health Dis.
- Chen, B. E., Kondo, M., Garnier, A., Watson, F. L., Püettmann-Holgado, R., Lamar, D. R., et al. (2006). The Molecular Diversity of Dscam Is Functionally Required for Neuronal Wiring Specificity in *Drosophila*. *Cell* 125, 607–620. doi:10.1016/j.cell.2006.03.034
- Chen, W. V., and Maniatis, T. (2013). Clustered Protocadherins. *Development* 140, 3297–3302. doi:10.1242/dev.090621
- Cheng, L. E., and Reed, R. R. (2007). Zfp423/OAZ Participates in a Developmental Switch during Olfactory Neurogenesis. *Neuron* 54, 547–557. doi:10.1016/j.neuron.2007.04.029
- Chesler, A. T., Zou, D.-J., Le Pichon, C. E., Peterlin, Z. A., Matthews, G. A., Pei, X., et al. (2007). A G protein/cAMP Signal cascade Is Required for Axonal Convergence into Olfactory Glomeruli. *Pnas* 104, 1039–1044. doi:10.1073/pnas.0609215104
- Chess, A., Simon, L., Cedar, H., and Axel, R. (1994). Allelic Inactivation Regulates Olfactory Receptor Gene Expression. *Cell* 78, 823–834. doi:10.1016/s0092-8674(94)90562-2
- Clowney, E. J., LeGros, M. A., Mosley, C. P., Clowney, F. G., Markenskoff-Papadimitriou, E. C., Myllys, M., et al. (2012). Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression. *Cell* 151, 724–737. doi:10.1016/j.cell.2012.09.043
- Collins, A. M., and Watson, C. T. (2018). Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire. *Front. Immunol.* 9, 2249. doi:10.3389/fimmu.2018.02249
- Cross, G. A. M., Kim, H.-S., and Wickstead, B. (2014). Capturing the Variant Surface Glycoprotein Repertoire (The VSGnome) of *Trypanosoma Brucei* Lister 427. *Mol. Biochem. Parasitol.* 195, 59–73. doi:10.1016/j.molbiopara.2014.06.004
- Crotty, S. (2014). T Follicular Helper Cell Differentiation, Function, and Roles in Disease. *Immunity* 41, 529–542. doi:10.1016/j.immuni.2014.10.004
- Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J., and Sanjuán, R. (2015). Extremely High Mutation Rate of HIV-1 *In Vivo*. *Plos Biol.* 13 (9), e1002251. doi:10.1371/journal.pbio.1002251
- da Silva, M. S., Hovel-Miner, G. A., Briggs, E. M., Elias, M. C., and McCulloch, R. (2018). Evaluation of Mechanisms that May Generate DNA Lesions Triggering Antigenic Variation in African Trypanosomes. *Plos Pathog.* 14, e1007321. doi:10.1371/journal.ppat.1007321
- Dalton, R. P., Lyons, D. B., and Lomvardas, S. (2013). Co-Opting the Unfolded Protein Response to Elicit Olfactory Receptor Feedback. *Cell* 155, 321–332. doi:10.1016/j.cell.2013.09.033
- Devlin, R., Marques, C. A., Paape, D., Prorocic, M., Zurita-Leal, A. C., Campbell, S. J., et al. (2016). Mapping Replication Dynamics in *Trypanosoma Brucei* Reveals a Link with Telomere Transcription and Antigenic Variation. *Elife* 5, e12765. doi:10.7554/eLife.12765
- Duveau, F., Hodgins-Davis, A., Metzger, B. P., Yang, B., Tryban, S., Walker, E. A., et al. (2018). Fitness Effects of Altering Gene Expression Noise in *Saccharomyces cerevisiae*. *Elife* 7, e37272. doi:10.7554/eLife.37272
- Eberle, A. B., Herrmann, K., Jäck, H.-M., and Mühlemann, O. (2009). Equal Transcription Rates of Productively and Nonproductively Rearranged Immunoglobulin μ Heavy Chain Alleles in a Pro-B Cell Line. *RNA* 15, 1021–1028. doi:10.1261/rna.1516409
- Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., et al. (2004). Mice Cloned from Olfactory Sensory Neurons. *Nature* 428, 44–49. doi:10.1038/nature02375
- Ersfeld, K., Melville, S. E., and Gull, K. (1999). Nuclear and Genome Organization of *Trypanosoma Brucei*. *Parasitol. Today* 15, 58–63. doi:10.1016/S0169-4758(98)01378-7
- Escobar, L. L., Hänisch, B., Halliday, C., Dean, S., Sunter, J. D., Wheeler, R. J., et al. (2021). *Monoallelic Antigen Expression in Trypanosomes Requires a Stage-specific Transcription Activator*. Headington, Oxford: Cold Spring Harbor Laboratory. doi:10.1101/2021.05.06.442931
- Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., et al. (2005). Monoallelic yet Combinatorial Expression of Variable Exons of the Protocadherin- α Gene Cluster in Single Neurons. *Nat. Genet.* 37, 171–176. doi:10.1038/ng1500
- Faria, J., Glover, L., Hutchinson, S., Boehm, C., Field, M. C., and Horn, D. (2019). Monoallelic Expression and Epigenetic Inheritance Sustained by a *Trypanosoma Brucei* Variant Surface Glycoprotein Exclusion Complex. *Nat. Commun.* 10 (1), 1–14. doi:10.1038/s41467-019-10823-8
- Faria, J., Luzak, V., Müller, L. S. M., Brink, B. G., Hutchinson, S., Glover, L., et al. (2021). Spatial Integration of Transcription and Splicing in a Dedicated Compartment Sustains Monogenic Antigen Expression in African Trypanosomes. *Nat. Microbiol.* 6 (3), 289–300. doi:10.1038/s41564-020-00833-4
- Feederle, R., and Schepers, A. (2017). Antibodies Specific for Nucleic Acid Modifications. *RNA Biol.* 14, 1089–1098. doi:10.1080/15476286.2017.1295905
- Ferreira, T., Wilson, S. R., Choi, Y. G., Rizzo, D., Dudoit, S., Speed, T. P., et al. (2014). Silencing of Odorant Receptor Genes by G Protein $\beta\gamma$ Signaling Ensures the Expression of One Odorant Receptor Per Olfactory Sensory Neuron. *Neuron* 81, 847–859. doi:10.1016/j.neuron.2014.01.001
- Figueiredo, L. M., Janzen, C. J., and Cross, G. A. M. (2008). A Histone Methyltransferase Modulates Antigenic Variation in African Trypanosomes. *Plos Biol.* 6, e161. doi:10.1371/journal.pbio.0060161
- Fleischmann, A., Shykind, B. M., Sosulski, D. L., Franks, K. M., Glinka, M. E., Mei, D. F., et al. (2008). Mice with a “Monoclonal Nose”: Perturbations in an Olfactory Map Impair Odor Discrimination. *Neuron* 60, 1068–1081. doi:10.1016/j.neuron.2008.10.046
- Forrest, S., and Oprea, M. (1999). *Antibody Repertoires and Pathogen Recognition: The Role of Germline Diversity and Somatic Hypermutation [WWW Document]*. Available at: <https://www.semanticscholar.org/paper/Antibody-repertoires-and-pathogen-recognition%3A-the-Forrest-Oprea/7be2f1a22f928d3a0f3ca8ca5129eef34987997> (accessed 1 12, 20).
- Galler, G. R., Mundt, C., Parker, M., Pelanda, R., Mårtensson, I.-L., and Winkler, T. H. (2004). Surface μ Heavy Chain Signals Down-Regulation of the V(D)J-Recombinase Machinery in the Absence of Surrogate Light Chain Components. *J. Exp. Med.* 199, 1523–1532. doi:10.1084/jem.20031523
- Garrett, A. M., Bosch, P. J., Steffen, D. M., Fuller, L. C., Marcucci, C. G., Koch, A. A., et al. (2019). CRISPR/Cas9 Interrogation of the Mouse *Pcdhg* Gene Cluster Reveals a Crucial Isoform-specific Role for *Pcdhg4*. *PLOS Genet.* 15, e1008554–32. doi:10.1371/journal.pgen.1008554
- Garrett, A. M., Khalil, A., Walton, D. O., and Burgess, R. W. (2018). DSCAM Promotes Self-Avoidance in the Developing Mouse Retina by Masking the Functions of Cadherin Superfamily Members. *Proc. Natl. Acad. Sci. USA* 115, E10216–E10224. doi:10.1073/pnas.1809430115
- Glover, L., Alsford, S., and Horn, D. (2013). DNA Break Site at Fragile Subtelomeres Determines Probability and Mechanism of Antigenic Variation in African Trypanosomes. *Plos Pathog.* 9, e1003260. doi:10.1371/journal.ppat.1003260
- Glover, L., and Horn, D. (2006). Repression of Polymerase I-mediated Gene Expression at *Trypanosoma Brucei* Telomeres. *EMBO Rep.* 7, 93–99. doi:10.1038/sj.embor.7400575
- Glover, L., Hutchinson, S., Alsford, S., and Horn, D. (2016). VEX1 Controls the Allelic Exclusion Required for Antigenic Variation in Trypanosomes. *Proc. Natl. Acad. Sci. USA* 113, 7225–7230. doi:10.1073/pnas.1600344113
- Godfrey, P. A., Malnic, B., and Buck, L. B. (2004). The Mouse Olfactory Receptor Gene Family. *Pnas* 101, 2156–2161. doi:10.1073/pnas.0308051100
- Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., et al. (2005). Epigenetic Ontogeny of the *Igk* Locus during B Cell Development. *Nat. Immunol.* 6, 198–203. doi:10.1038/ni1154
- Goodman, K. M., Rubinstein, R., Dan, H., Bahna, F., Mannepalli, S., Ahlsén, G., et al. (2017). Protocadherin Cis-Dimer Architecture and Recognition Unit Diversity. *Proc. Natl. Acad. Sci. USA* 114, E9829–E9837. doi:10.1073/pnas.1713449114
- Graveley, B. R. (2005). Mutually Exclusive Splicing of the Insect *Dscam* Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. *Cell* 123, 65–73. doi:10.1016/j.cell.2005.07.028

- Grawunder, U., Leu, T. M. J., Schatz, D. G., Werner, A., Rolink, A. G., Melchers, F., et al. (1995). Down-regulation of RAG1 and RAG2 Gene Expression in preB Cells after Functional Immunoglobulin Heavy Chain Rearrangement. *Immunity* 3, 601–608. doi:10.1016/1074-7613(95)90131-0
- Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K. E., Li, W., et al. (2012). CTCF/cohesin-mediated DNA Looping Is Required for Protocadherin Promoter Choice. *Proc. Natl. Acad. Sci.* 109, 21081–21086. doi:10.1073/pnas.1219280110
- Hall, J. P. J., Wang, H., and Barry, J. D. (2013). Mosaic VSGs and the Scale of Trypanosoma Brucei Antigenic Variation. *Plos Pathog.* 9, e1003502. doi:10.1371/journal.ppat.1003502
- Hanchate, N. K., Kondoh, K., Lu, Z., Kuang, D., Ye, X., Qiu, X., et al. (2015). Single-cell Transcriptomics Reveals Receptor Transformations during Olfactory Neurogenesis. *Science* 350, 1251–1255. doi:10.1126/science.aad2456
- Hartley, C. L., and McCulloch, R. (2008). Trypanosoma Brucei BRCA2 Acts in Antigenic Variation and Has Undergone a Recent Expansion in BRC Repeat Number that Is Important during Homologous Recombination. *Mol. Microbiol.* 68, 1237–1251. doi:10.1111/j.1365-2958.2008.06230.x
- Hasegawa, S., Kumagai, M., Hagihara, M., Nishimaru, H., Hirano, K., Kaneko, R., et al. (2016). Distinct and Cooperative Functions for the Protocadherin- α , - β and - γ Clusters in Neuronal Survival and Axon Targeting. *Front. Mol. Neurosci.* 9, 155. doi:10.3389/fnfmol.2016.00155
- Hattori, D., Chen, Y., Matthews, B. J., Salwinski, L., Sabatti, C., Grueber, W. B., et al. (2009). Robust Discrimination between Self and Non-self Neurites Requires Thousands of Dscam1 Isoforms. *Nature* 461, 644–648. doi:10.1038/nature08431
- Hattori, D., Demir, E., Kim, H. W., Viragh, E., Zipursky, S. L., and Dickson, B. J. (2007). Dscam Diversity Is Essential for Neuronal Wiring and Self-Recognition. *Nature* 449, 223–227. doi:10.1038/nature06099
- Hausmann, I. U., Ustaoglu, P., Brauer, U., Hemani, Y., Dix, T. C., and Soller, M. (2019). Plasmid-based gap-repair Recombineered Transgenes Reveal a central Role for Introns in Mutually Exclusive Alternative Splicing in Down Syndrome Cell Adhesion Molecule Exon 4. *Nucleic Acids Res.* 47, 1389–1403. doi:10.1093/nar/gky1254
- Hemani, Y., and Soller, M. (2012). Mechanisms of Drosophila Dscam Mutually Exclusive Splicing Regulation. *Biochem. Soc. Trans.* 40, 804–809. doi:10.1042/BST20120060
- Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., et al. (2008). Telomeric Expression Sites Are Highly Conserved in Trypanosoma Brucei. *PLOS ONE* 3, e3527. doi:10.1371/journal.pone.0003527
- Hetz, C., Zhang, K., and Kaufman, R. J. (2020). Mechanisms, Regulation and Functions of the Unfolded Protein Response. *Nat. Rev. Mol. Cell Biol.* 21, 421–438. doi:10.1038/s41580-020-0250-z
- Hewitt, S. L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., et al. (2009). RAG-1 and ATM Coordinate Monoallelic Recombination and Nuclear Positioning of Immunoglobulin Loci. *Nat. Immunol.* 10, 655–664. doi:10.1038/ni.1735
- Hing, H., Xiao, J., Harden, N., Lim, L., and Zipursky, S. L. (1999). Pak Functions Downstream of Dock to Regulate Photoreceptor Axon Guidance in Drosophila. *Cell* 97, 853–863. doi:10.1016/S0092-8674(00)80798-9
- Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T., and Yagi, T. (2012). Single-neuron Diversity Generated by Protocadherin- β Cluster in Mouse central and Peripheral Nervous Systems. *Front. Mol. Neurosci.* 5, 90. doi:10.3389/fnfmol.2012.00090
- Hirayama, T., and Yagi, T. (2017). Regulation of Clustered Protocadherin Genes in Individual Neurons. *Semin. Cell Developmental Biol.* 69, 122–130. doi:10.1016/j.semcdb.2017.05.026
- Hirota, J., and Mombaerts, P. (2004). The LIM-Homeodomain Protein Lhx2 Is Required for Complete Development of Mouse Olfactory Sensory Neurons. *Proc. Natl. Acad. Sci.* 101, 8751–8755. doi:10.1073/pnas.0400940101
- Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P., and Cross, G. A. M. (1980). Novel Expression-Linked Copies of the Genes for Variant Surface Antigens in Trypanosomes. *Nature* 284, 78–80. doi:10.1038/284078a0
- Horn, D. (2014). Antigenic Variation in African Trypanosomes. *Mol. Biochem. Parasitol.* 195, 123–129. doi:10.1016/j.molbiopara.2014.05.001
- Horn, D., and Cross, G. A. M. (1997). Analysis of Trypanosoma Brucei Vsg Expression Site Switching *In Vitro*. *Mol. Biochem. Parasitol.* 84, 189–201. doi:10.1016/s0166-6851(96)02794-6
- Horn, D., and McCulloch, R. (2010). Molecular Mechanisms Underlying the Control of Antigenic Variation in African Trypanosomes. *Curr. Opin. Microbiol.* 13, 700–705. doi:10.1016/j.mib.2010.08.009
- Hovel-Miner, G. A., Boothroyd, C. E., Mugnier, M., Dreesen, O., Cross, G. A. M., and Papavasiliou, F. N. (2012). Telomere Length Affects the Frequency and Mechanism of Antigenic Variation in Trypanosoma Brucei. *Plos Pathog.* 8, e1002900. doi:10.1371/journal.ppat.1002900
- Hughes, M. E., Bortnick, R., Tsubouchi, A., Bäumer, P., Kondo, M., Uemura, T., et al. (2007). Homophilic Dscam Interactions Control Complex Dendrite Morphogenesis. *Neuron* 54, 417–427. doi:10.1016/j.neuron.2007.04.013
- Hummel, T., Vasconcelos, M. L., Clemens, J. C., Fishilevich, Y., Vossball, L. B., and Zipursky, S. L. (2003). Axonal Targeting of Olfactory Receptor Neurons in Drosophila Is Controlled by Dscam. *Neuron* 37, 221–231. doi:10.1016/s0896-6273(02)01183-2
- Hutchinson, S., Foulon, S., Crouzols, A., Menafra, R., Rotureau, B., Griffiths, A. D., et al. (2021). The Establishment of Variant Surface Glycoprotein Monoallelic Expression Revealed by Single-Cell RNA-Seq of Trypanosoma Brucei in the Tsetse Fly Salivary Glands. *BioRxiv* 1, 1. doi:10.1101/2021.03.01.433049
- Iakovlev, M., Faravelli, S., and Becskei, A. (2021). Gene Families with Stochastic Exclusive Gene Choice Underlie Cell Adhesion in Mammalian Cells. *Front. Cell. Dev. Biol.* 9. doi:10.3389/fcell.2021.642212
- Imai, T., Suzuki, M., and Sakano, H. (2006). Odorant Receptor-Derived cAMP Signals Direct Axonal Targeting. *Science* 314, 657–661. doi:10.1126/science.1131794
- Ing-Esteves, S., Kostadinov, D., Marocha, J., Sing, A. D., Joseph, K. S., Laboulaye, M. A., et al. (2018). Combinatorial Effects of Alpha- and Gamma-Protocadherins on Neuronal Survival and Dendritic Self-Avoidance. *J. Neurosci.* 38, 2713–2729. doi:10.1523/JNEUROSCI.3035-17.2018
- Jacob, J., Przylepa, J., Miller, C., and Kelsoe, G. (1993). *In Situ* studies of the Primary Immune Response to (4-Hydroxy-3-Nitrophenyl)acetyl. III. The Kinetics of V Region Mutation and Selection in Germinal center B Cells. *J. Exp. Med.* 178, 1293–1307. doi:10.1084/jem.178.4.1293
- Jayaraman, S., Harris, C., Paxton, E., Donachie, A.-M., Vaikkinen, H., McCulloch, R., et al. (2019). Application of Long Read Sequencing to Determine Expressed Antigen Diversity in Trypanosoma Brucei Infections. *Plos Negl. Trop. Dis.* 13, e0007262. doi:10.1371/journal.pntd.0007262
- Jin, Y., Dong, H., Shi, Y., and Bian, L. (2018). Mutually Exclusive Alternative Splicing of Pre-mRNAs. *WIREs RNA* 9, e1468. doi:10.1002/wrna.1468
- Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F. W. (2006). Mechanism and Control of V(D)J Recombination at the Immunoglobulin Heavy Chain Locus. *Annu. Rev. Immunol.* 24, 541–570. doi:10.1146/annurev.immunol.23.021704.115830
- Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., et al. (2006). Allelic Gene Regulation of Pcdh- α and Pcdh- γ Clusters Involving Both Monoallelic and Biallelic Expression in Single Purkinje Cells. *J. Biol. Chem.* 281, 30551–30560. doi:10.1074/jbc.M605677200
- Kassem, A., Pays, E., and Vanhamme, L. (2014). Transcription Is Initiated on Silent Variant Surface Glycoprotein Expression Sites Despite Monoallelic Expression in Trypanosoma Brucei. *Proc. Natl. Acad. Sci.* 111, 8943–8948. doi:10.1073/pnas.1404873111
- Kawahara, T., Siegel, T. N., Ingram, A. K., Alsford, S., Cross, G. A. M., and Horn, D. (2008). Two Essential MYST-Family Proteins Display Distinct Roles in Histone H4K10 Acetylation and Telomeric Silencing in Trypanosomes. *Mol. Microbiol.* 69, 1054–1068. doi:10.1111/j.1365-2958.2008.06346.x
- Kehayova, P., Monahan, K., Chen, W., and Maniatis, T. (2011). Regulatory Elements Required for the Activation and Repression of the Protocadherin-Gene Cluster. *Proc. Natl. Acad. Sci.* 108, 17195–17200. doi:10.1073/pnas.1114357108
- Khamlichi, A. A., and Feil, R. (2018). Parallels between Mammalian Mechanisms of Monoallelic Gene Expression. *Trends Genet.* 34 (12), 954–971. doi:10.1016/j.tig.2018.08.005
- Khan, M., Vaes, E., and Mombaerts, P. (2011). Regulation of the Probability of Mouse Odorant Receptor Gene Choice. *Cell* 147, 907–921. doi:10.1016/j.cell.2011.09.049
- Kim, H.-S., and Cross, G. A. M. (2011). Identification of Trypanosoma Brucei RMI1/BLAP75 Homologue and its Roles in Antigenic Variation. *PLoS One* 6, e25313. doi:10.1371/journal.pone.0025313

- Kim, H.-S., and Cross, G. A. M. (2010). TOPO3 α Influences Antigenic Variation by Monitoring Expression-Site-Associated VSG Switching in *Trypanosoma Brucei*. *Plos Pathog.* 6, e1000992. doi:10.1371/journal.ppat.1000992
- Kirkham, J. K., Park, S. H., Nguyen, T. N., Lee, J. H., and Günzl, A. (2016). Dynein Light Chain LC8 Is Required for RNA Polymerase I-Mediated Transcription in *Trypanosoma Brucei*, Facilitating Assembly and Promoter Binding of Class I Transcription Factor A. *Mol. Cel. Biol.* 36, 95–107. doi:10.1128/MCB.00705-15
- Klein, L., Kyewski, B., Allen, P. M., and Hogquist, K. A. (2014). Positive and Negative Selection of the T Cell Repertoire: what Thymocytes See (And Don't See). *Nat. Rev. Immunol.* 14, 377–391. doi:10.1038/nri3667
- Lefebvre, J. L., Kostadinov, D., Chen, W. V., Maniatis, T., and Sanes, J. R. (2012). Protocadherins Mediate Dendritic Self-Avoidance in the Mammalian Nervous System. *Nature* 488, 517–521. doi:10.1038/nature11305
- Lewcock, J. W., and Reed, R. R. (2004). A Feedback Mechanism Regulates Monoallelic Odorant Receptor Expression. *Proc. Natl. Acad. Sci.* 101, 1069–1074. doi:10.1073/pnas.0307986100
- Le Gros, M. A., Clowney, E. J., Magklara, A., Yen, A., Markenscoff-Papadimitriou, E., Colquitt, B., et al. (2016). Soft X-Ray Tomography Reveals Gradual Chromatin Compaction and Reorganization during Neurogenesis *In Vivo*. *Cel. Rep.* 17, 2125–2136. doi:10.1016/j.celrep.2016.10.060
- Li, B. (2015). DNA Double-Strand Breaks and Telomeres Play Important Roles in *Trypanosoma Brucei* Antigenic Variation. *Eukaryot. Cel.* 14, 196–205. doi:10.1128/EC.00207-14
- Li, F., and Eckhardt, L. A. (2009). A Role for the IgH Intronic Enhancer E μ in Enforcing Allelic Exclusion. *J. Exp. Med.* 206, 153–167. doi:10.1084/jem.20081202
- Li, J., Ishii, T., Feinstein, P., and Mombaerts, P. (2004). Odorant Receptor Gene Choice Is Reset by Nuclear Transfer from Mouse Olfactory Sensory Neurons. *Nature* 428, 393–399. doi:10.1038/nature02433
- Liang, H.-E., Hsu, L.-Y., Cado, D., and Schissel, M. S. (2004). Variegated Transcriptional Activation of the Immunoglobulin κ Locus in Pre-B Cells Contributes to the Allelic Exclusion of Light-Chain Expression. *Cell* 118, 19–29. doi:10.1016/j.cell.2004.06.019
- Little, A. J., Matthews, A., Oettinger, M., Roth, D. B., and Schatz, D. G. (2015). “The Mechanism of V(D)J Recombination,” in *Molecular Biology of B Cells*. Editors F. W. Alt, T. Honjo, A. Radbruch, and M. Reth. Second Edition (London: Academic Press), 13–34. doi:10.1016/B978-0-12-397933-9.00002-3
- Liu, A. Y. C., Van der Ploeg, L. H. T., Rijsewijk, F. A. M., Borst, P., and Chambon, P. (1983). The Transposition Unit of Variant Surface Glycoprotein Gene 118 of *Trypanosoma Brucei*. *J. Mol. Biol.* 167, 57–75. doi:10.1016/s0022-2836(83)80034-5
- Liu, H., Pizzano, S., Li, R., Zhao, W., Veling, M. W., Hu, Y., et al. (2020). isoTarget: A Genetic Method for Analyzing the Functional Diversity of Splicing Isoforms *In Vivo*. *Cel. Rep.* 33, 108361. doi:10.1016/j.celrep.2020.108361
- López-Farfán, D., Bart, J.-M., Rojas-Barros, D. I., and Navarro, M. (2014). SUMOylation by the E3 Ligase TbSIZ1/PIAS1 Positively Regulates VSG Expression in *Trypanosoma Brucei*. *Plos Pathog.* 10, e1004545. doi:10.1371/journal.ppat.1004545
- Lyons, D. B., Allen, W. E., Goh, T., Tsai, L., Barnea, G., and Lomvardas, S. (2013). An Epigenetic Trap Stabilizes Singular Olfactory Receptor Expression. *Cell* 154, 325–336. doi:10.1016/j.cell.2013.06.039
- Lyons, D. B., Magklara, A., Goh, T., Sampath, S. C., Schaefer, A., Schotta, G., et al. (2014). Heterochromatin-mediated Gene Silencing Facilitates the Diversification of Olfactory Neurons. *Cel. Rep.* 9, 884–892. doi:10.1016/j.celrep.2014.10.001
- Machado, C. F., Nagai, M. H., Lyra, C. S., Reis-Silva, T. M., Xavier, A. M., Glezer, I., et al. (2017). Conditional Deletion of Ric-8b in Olfactory Sensory Neurons Leads to Olfactory Impairment. *J. Neurosci.* 37, 12202–12213. doi:10.1523/JNEUROSCI.0943-17.2017
- Magklara, A., and Lomvardas, S. (2013). Stochastic Gene Expression in Mammals: Lessons from Olfaction. *Trends Cell Biology* 23 (9), 449–456. doi:10.1016/j.tcb.2013.04.005
- Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., et al. (2011). An Epigenetic Signature for Monoallelic Olfactory Receptor Expression. *Cell* 145, 555–570. doi:10.1016/j.cell.2011.03.040
- Mah, K. M., and Weiner, J. A. (2016). “Clustered Protocadherins,” in *The Cadherin Superfamily: Key Regulators of Animal Development and Physiology*. Editors S. T. Suzuki and S. Hirano (Tokyo: Springer Japan), 195–221. doi:10.1007/978-4-431-56033-3_8
- Malnic, B., Godfrey, P. A., and Buck, L. B. (2004). The Human Olfactory Receptor Gene Family. *Proc. Natl. Acad. Sci.* 101, 2584–2589. doi:10.1073/pnas.0307882100
- Malnic, B., Hirono, J., Sato, T., and Buck, L. B. (1999). Combinatorial Receptor Codes for Odors. *Cell* 96, 713–723. doi:10.1016/S0092-8674(00)80581-4
- Mancini, M., Bassani, S., and Passafaro, M. (2020). Right Place at the Right Time: How Changes in Protocadherins Affect Synaptic Connections Contributing to the Etiology of Neurodevelopmental Disorders. *Cells* 9, 2711. doi:10.3390/cells9122711
- Markenscoff-Papadimitriou, E., Allen, W. E., Colquitt, B. M., Goh, T., Murphy, K. K., Monahan, K., et al. (2014). Enhancer Interaction Networks as a Means for Singular Olfactory Receptor Expression. *Cell* 159, 543–557. doi:10.1016/j.cell.2014.09.033
- Massah, S., Beischlag, T. V., and Prefontaine, G. G. (2015). Epigenetic Events Regulating Monoallelic Gene Expression. *Crit. Rev. Biochem. Mol. Biol.* 50, 337–358. doi:10.3109/10409238.2015.1064350
- Matthews, B. J., Kim, M. E., Flanagan, J. J., Hattori, D., Clemens, J. C., Zipursky, S. L., et al. (2007). Dendrite Self-Avoidance Is Controlled by Dscam. *Cell* 129, 593–604. doi:10.1016/j.cell.2007.04.013
- Matthews, K. R. (2005). The Developmental Cell Biology of *Trypanosoma Brucei*. *J. Cel. Sci.* 118, 283–290. doi:10.1242/jcs.01649
- May, G. E., Olson, S., McManus, C. J., and Graveley, B. R. (2011). Competing RNA Secondary Structures Are Required for Mutually Exclusive Splicing of the Dscam Exon 6 Cluster. *Rna* 17, 222–229. doi:10.1261/rna.2521311
- McCulloch, R., and Barry, J. D. (1999). A Role for RAD51 and Homologous Recombination in *Trypanosoma Brucei* Antigenic Variation. *Genes Development* 13, 2875–2888. doi:10.1101/gad.13.21.2875
- McCulloch, R., Cobbold, C. A., Figueiredo, L., Jackson, A., Morrison, L. J., Mugnier, M. R., et al. (2017). Emerging Challenges in Understanding *Trypanosoma* Antigenic Variation. *Emerg. Top. Life Sci.* 1, 585–592. doi:10.1042/ETLS20170104
- McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W., and Weigert, M. (1984). Generation of Antibody Diversity in the Immune Response of BALB/c Mice to Influenza Virus Hemagglutinin. *Proc. Natl. Acad. Sci.* 81, 3180–3184. doi:10.1073/pnas.81.10.3180
- Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J.-h., Walz, T., Wang, J.-h., et al. (2007). Structural Basis of Dscam Isoform Specificity. *Nature* 449, 487–491. doi:10.1038/nature06147
- Metzger, B. P. H., Yuan, D. C., Gruber, J. D., Duvéau, F., and Wittkopp, P. J. (2015). Selection on Noise Constrains Variation in a Eukaryotic Promoter. *Nature* 521 (7552), 344–347. doi:10.1038/nature14244
- Millard, S. S., Lu, Z., Zipursky, S. L., and Meinertzhagen, I. A. (2010). Drosophila Dscam Proteins Regulate Postsynaptic Specificity at Multiple-Contact Synapses. *Neuron* 67, 761–768. doi:10.1016/j.neuron.2010.08.030
- Miura, S. K., Martins, A., Zhang, K. X., Graveley, B. R., and Zipursky, S. L. (2013). Probabilistic Splicing of Dscam1 Establishes Identity at the Level of Single Neurons. *Cell* 155, 1166–1177. doi:10.1016/j.cell.2013.10.018
- Molmby, M. J., Keeler, A. B., and Weiner, J. A. (2016). Homophilic Protocadherin Cell-Cell Interactions Promote Dendrite Complexity. *Cel. Rep.* 15, 1037–1050. doi:10.1016/j.celrep.2016.03.093
- Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-Mediated Trans Interactions Regulate Olfactory Receptor Choice. *Nature* 565, 448–453. doi:10.1038/s41586-018-0845-0
- Monahan, K., and Lomvardas, S. (2015). Monoallelic Expression of Olfactory Receptors. *Annu. Rev. Cel. Dev. Biol.* 31, 721–740. doi:10.1146/annurev-cellbio-100814-125308
- Monahan, K., Rudnick, N. D., Kehayova, P. D., Pauli, F., Newberry, K. M., Myers, R. M., et al. (2012). Role of CCCTC Binding Factor (CTCF) and Cohesin in the Generation of Single-Cell Diversity of Protocadherin- Gene Expression. *Proc. Natl. Acad. Sci.* 109, 9125–9130. doi:10.1073/pnas.1205074109
- Monahan, K., Schieren, I., Cheung, J., Mumbey-Wafula, A., Monuki, E. S., and Lomvardas, S. (2017). Cooperative Interactions Enable Singular Olfactory Receptor Expression in Mouse Olfactory Neurons. *eLife* 6. doi:10.7554/eLife.28620

- Morishita, H., and Yagi, T. (2007). Protocadherin Family: Diversity, Structure, and Function. *Curr. Opin. Cel. Biol.* 19, 584–592. doi:10.1016/j.ceb.2007.09.006
- Morrison, L. J., Majiwa, P., Read, A. F., and Barry, J. D. (2005). Probabilistic Order in Antigenic Variation of Trypanosoma Brucei. *Int. J. Parasitol.* 35, 961–972. doi:10.1016/j.ijpara.2005.05.004
- Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., et al. (2001). Asynchronous Replication and Allelic Exclusion in the Immune System. *Nature* 414, 221–225. doi:10.1038/35102606
- Mountoufaris, G., Chen, W. V., Hirabayashi, Y., O’Keeffe, S., Chevee, M., Nwakeze, C. L., et al. (2017). Multiclusted Pcdh Diversity Is Required for Mouse Olfactory Neural Circuit Assembly. *Science* 356, 411–414. doi:10.1126/science.aai8801
- Mugnier, M. R., Cross, G. A. M., and Papavasiliou, F. N. (2015). The *In Vivo* Dynamics of Antigenic Variation in Trypanosoma Brucei. *Science* 347, 1470–1473. doi:10.1126/science.aaa4502
- Mugnier, M. R., Stebbins, C. E., and Papavasiliou, F. N. (2016). Masters of Disguise: Antigenic Variation and the VSG Coat in Trypanosoma Brucei. *Plos Pathog.* 12, e1005784. doi:10.1371/journal.ppat.1005784
- Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., et al. (2018). Genome Organization and DNA Accessibility Control Antigenic Variation in Trypanosomes. *Nature* 563, 121–125. doi:10.1038/s41586-018-0619-8
- Nagai, M. H., Armelin-Correa, L. M., and Malnic, B. (2016). Monogenic and Monoallelic Expression of Odorant Receptors. *Mol. Pharmacol.* 90, 633–639. doi:10.1124/mol.116.104745
- Narayanan, M. S., Kushwaha, M., Ersfeld, K., Fullbrook, A., Stanne, T. M., and Rudenko, G. (2011). NLP Is a Novel Transcription Regulator Involved in VSG Expression Site Control in Trypanosoma Brucei. *Nucleic Acids Res.* 39, 2018–2031. doi:10.1093/nar/gkq950
- Navarro, M., and Gull, K. (2001). A Pol I Transcriptional Body Associated with VSG Mono-Allelic Expression in Trypanosoma Brucei. *Nature* 414, 759–763. doi:10.1038/414759a
- Nemazee, D. (2017). Mechanisms of central Tolerance for B Cells. *Nat. Rev. Immunol.* 17, 281–294. doi:10.1038/nri.2017.19
- Neves, G., Zucker, J., Daly, M., and Chess, A. (2004). Stochastic yet Biased Expression of Multiple Dscam Splice Variants by Individual Cells. *Nat. Genet.* 36, 240–246. doi:10.1038/ng1299
- Nguyen, M. Q., Zhou, Z., Marks, C. A., Ryba, N. J. P., and Belluscio, L. (2007). Prominent Roles for Odorant Receptor Coding Sequences in Allelic Exclusion. *Cell* 131, 1009–1017. doi:10.1016/j.cell.2007.10.050
- Nguyen, T. N., Müller, L. S. M., Park, S. H., Siegel, T. N., and Günzl, A. (2014). Promoter Occupancy of the Basal Class I Transcription Factor A Differs Strongly between Active and Silent VSG Expression Sites in Trypanosoma Brucei. *Nucleic Acids Res.* 42, 3164–3176. doi:10.1093/nar/gkt1301
- Nguyen, T. N., Nguyen, B. N., Lee, J. H., Panigrahi, A. K., and Günzl, A. (2012). Characterization of a Novel Class I Transcription Factor A (CITFA) Subunit that Is Indispensable for Transcription by the Multifunctional RNA Polymerase I of Trypanosoma Brucei. *Eukaryot. Cel.* 11, 1573–1581. doi:10.1128/EC.00250-12
- Nishizumi, H., Kumasaka, K., Inoue, N., Nakashima, A., and Sakano, H. (2007). Deletion of the Core-H Region in Mice Abolishes the Expression of Three Proximal Odorant Receptor Genes in Cis. *Proc. Natl. Acad. Sci.* 104, 20067–20072. doi:10.1073/pnas.0706544105
- Nozawa, M., and Nei, M. (2009). Genomic Drift and Copy Number Variation of Chemosensory Receptor Genes in Humans and Mice. *Cytogenet. Genome Res.* 123, 263–269. doi:10.1159/000184716
- Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, Adjacent Genes that Synergistically Activate V(D)J Recombination. *Science* 248, 1517–1523. doi:10.1126/science.2360047
- Okada, T., Miller, M. J., Parker, I., Krummel, M. F., Neighbors, M., Hartley, S. B., et al. (2005). Antigen-engaged B Cells Undergo Chemotaxis toward the T Zone and Form Motile Conjugates with Helper T Cells. *Plos Biol.* 3, e150. doi:10.1371/journal.pbio.0030150
- Olender, T., Lancet, D., and Nebert, D. W. (2008). Update on the Olfactory Receptor (OR) Gene Superfamily. *Hum. Genomics* 3, 87. doi:10.1186/1479-7364-3-1-87
- Olson, S., Blanchette, M., Park, J., Savva, Y., Yeo, G. W., Yeakley, J. M., et al. (2007). A Regulator of Dscam Mutually Exclusive Splicing Fidelity. *Nat. Struct. Mol. Biol.* 14, 1134–1140. doi:10.1038/nsmb1339
- Outters, P., Jaeger, S., Zaarour, N., and Ferrier, P. (2015). “Long-Range Control of V(D)J Recombination & Allelic Exclusion,” in *Advances in Immunology, Molecular Mechanisms that Orchestrate the Assembly of Antigen Receptor Loci*. Editor C. Murre (Academic Press), 363–413. doi:10.1016/bs.ai.2015.08.002
- Papavasiliou, F. N., and Schatz, D. G. (2000). Cell-cycle-regulated DNA Double-Strand Breaks in Somatic Hypermutation of Immunoglobulin Genes. *Nature* 408, 216–221. doi:10.1038/35041599
- Park, J. W., Parisky, K., Celotto, A. M., Reenan, R. A., and Graveley, B. R. (2004). Identification of Alternative Splicing Regulators by RNA Interference in Drosophila. *Proc. Natl. Acad. Sci.* 101, 15974–15979. doi:10.1073/pnas.0407004101
- Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Pérez-Morga, D. (2001). The VSG Expression Sites of Trypanosoma Brucei: Multipurpose Tools for the Adaptation of the Parasite to Mammalian Hosts. *Mol. Biochem. Parasitol.* 114, 1–16. doi:10.1016/S0166-6851(01)00242-0
- Perry, R. P., Kelley, D. E., Coleclough, C., Seidman, J. G., Leder, P., Tonegawa, S., et al. (1980). Transcription of Mouse Kappa Chain Genes: Implications for Allelic Exclusion. *Proc. Natl. Acad. Sci.* 77, 1937–1941. doi:10.1073/pnas.77.4.1937
- P. Ferrier (Editor) (2009). *V(D)J Recombination, Advances in Experimental Medicine and Biology* (New York: Springer-Verlag). doi:10.1007/978-1-4419-0296-2
- Pilzecker, B., and Jacobs, H. (2019). Mutating for Good: DNA Damage Responses during Somatic Hypermutation. *Front. Immunol.* 10. doi:10.3389/fimmu.2019.00438
- Povelones, M. L., Gluenz, E., Dembek, M., Gull, K., and Rudenko, G. (2012). Histone H1 Plays a Role in Heterochromatin Formation and VSG Expression Site Silencing in Trypanosoma Brucei. *Plos Pathog.* 8, e1003010. doi:10.1371/journal.ppat.1003010
- Raj, A., Rifkin, S. A., Andersen, E., and Van Oudenaarden, A. (2010). Variability in Gene Expression Underlies Incomplete Penetrance. *Nature* 463 (7283), 913–918. doi:10.1038/nature08781
- Reynolds, D., Hofmeister, B. T., Cliffe, L., Alabady, M., Siegel, T. N., Schmitz, R. J., et al. (2016). Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in Trypanosoma Brucei. *Plos Genet.* 12, e1005758. doi:10.1371/journal.pgen.1005758
- Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of Long-Range Regulatory Elements in the Protocadherin- Gene Cluster. *Proc. Natl. Acad. Sci.* 103, 19719–19724. doi:10.1073/pnas.0609445104
- Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988). The Accuracy of Reverse Transcriptase from HIV-1. *Science* 242 (4882), 1171–1173. doi:10.1126/science.2460925
- Robin, J. D., Ludlow, A. T., Batten, K., Magdinier, F., Stadler, G., Wagner, K. R., et al. (2014). Telomere Position Effect: Regulation of Gene Expression with Progressive Telomere Shortening over Long Distances. *Genes Dev.* 28, 2464–2476. doi:10.1101/gad.251041.114
- Robinson, N. P., Burman, N., Melville, S. E., and Barry, J. D. (1999). Predominance of Duplicative VSG Gene Conversion in Antigenic Variation in African Trypanosomes. *Mol. Cel. Biol.* 19, 5839–5846. doi:10.1128/MCB.19.9.5839
- Roby, Y. A., Bushey, M. A., Cheng, L. E., Kulaga, H. M., Lee, S.-J., and Reed, R. R. (2012). Zfp423/OAZ Mutation Reveals the Importance of Olf/EBF Transcription Activity in Olfactory Neuronal Maturation. *J. Neurosci.* 32, 13679–13688a. doi:10.1523/JNEUROSCI.6190-11.2012
- Rodriguez, O. L., Gibson, W. S., Parks, T., Emery, M., Powell, J., Strahl, M., et al. (2020). A Novel Framework for Characterizing Genomic Haplotype Diversity in the Human Immunoglobulin Heavy Chain Locus. *Front. Immunol.* 11. doi:10.3389/fimmu.2020.02136
- Ron, D., and Walter, P. (2007). Signal Integration in the Endoplasmic Reticulum Unfolded Protein Response. *Nat. Rev. Mol. Cel. Biol.* 8, 519–529. doi:10.1038/nrm2199
- Rose, N. R. (2017). Negative Selection, Epitope Mimicry and Autoimmunity. *Curr. Opin. Immunol.* 49, 51–55. doi:10.1016/j.coi.2017.08.014

- Rothman, A., Feinstein, P., Hirota, J., and Mombaerts, P. (2005). The Promoter of the Mouse Odorant Receptor Gene M71. *Mol. Cell Neurosci.* 28, 535–546. doi:10.1016/j.mcn.2004.11.006
- Roy, A. L., Sen, R., and Roeder, R. G. (2011). Enhancer-promoter Communication and Transcriptional Regulation of *Igh*. *Trends Immunol.* 32, 532–539. doi:10.1016/j.it.2011.06.012
- Rudenko, G., McCulloch, R., Dirks-Mulder, A., and Borst, P. (1996). Telomere Exchange Can Be an Important Mechanism of Variant Surface Glycoprotein Gene Switching in *Trypanosoma Brucei*. *Mol. Biochem. Parasitol.* 80, 65–75. doi:10.1016/0166-6851(96)02669-2
- Sachse, S. M., Lievens, S., Ribeiro, L. F., Dascenco, D., Masschaele, D., Horré, K., et al. (2019). Nuclear Import of the DSCAM -cytoplasmic Domain Drives Signaling Capable of Inhibiting Synapse Formation. *EMBO J.* 38, e99669. doi:10.15252/embj.201899669
- Saura, A., Iribarren, P. A., Rojas-Barros, D., Bart, J. M., López-Farfán, D., Andrés-León, E., et al. (2019). SUMOylated SNF2PH Promotes Variant Surface Glycoprotein Expression in Bloodstream Trypanosomes. *EMBO Rep.* 20, e48029. doi:10.15252/embr.201948029
- Sawaya, M. R., Wojtowicz, W. M., Andre, I., Qian, B., Wu, W., Baker, D., et al. (2008). A Double S Shape Provides the Structural Basis for the Extraordinary Binding Specificity of Dscam Isoforms. *Cell* 134, 1007–1018. doi:10.1016/j.cell.2008.07.042
- Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). The V(D)J Recombination Activating Gene, RAG-1. *Cell* 59, 1035–1048. doi:10.1016/0092-8674(89)90760-5
- Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: Complex Genes in Simple Animals, Complex Animals yet Simple Genes. *Genes Dev.* 23, 147–156. doi:10.1101/gad.1752909
- Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., et al. (2000). Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell* 101, 671–684. doi:10.1016/S0092-8674(00)80878-8
- Schreiner, D., and Weiner, J. A. (2010). Combinatorial Homophilic Interaction between -protocadherin Multimers Greatly Expands the Molecular Diversity of Cell Adhesion. *Proc. Natl. Acad. Sci.* 107, 14893–14898. doi:10.1073/pnas.1004526107
- Schroeder, H. W., and Cavacini, L. (2010). Structure and Function of Immunoglobulins. *J. Allergy Clin. Immunol.* 125, S41–S52. doi:10.1016/j.jaci.2009.09.046
- Schulz, D., and Papavasiliou, F. N. (2016). The VEXing Problem of Monoallelic Expression in the African Trypanosome. *Proc. Natl. Acad. Sci. USA* 113, 7017–7019. doi:10.1073/pnas.1608546113
- Schulz, D., Zaringhalam, M., Papavasiliou, F. N., and Kim, H.-S. (2016). Base J and H3.V Regulate Transcriptional Termination in *Trypanosoma Brucei*. *Plos Genet.* 12, e1005762. doi:10.1371/journal.pgen.1005762
- Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., et al. (2003). Negative Feedback Regulation Ensures the One Receptor-One Olfactory Neuron Rule in Mouse. *Science* 302, 2088–2094. doi:10.1126/science.1089122
- Serwe, M., and Sablitzky, F. (1993). V(D)J Recombination in B Cells Is Impaired but Not Blocked by Targeted Deletion of the Immunoglobulin Heavy Chain Intron Enhancer. *EMBO J.* 12, 2321–2327. doi:10.1002/j.1460-2075.1993.tb05886.x
- Sharma, R., Ishimaru, Y., Davison, I., Ikegami, K., Chien, M.-S., You, H., et al. (2017). Olfactory Receptor Accessory Proteins Play Crucial Roles in Receptor Function and Gene Choice. *eLife* 6, e21895. doi:10.7554/eLife.21895
- Shykind, B. M. (2005). Regulation of Odorant Receptors: One Allele at a Time. *Hum. Mol. Genet.* 14, R33–R39. doi:10.1093/hmg/ddi105
- Sima, N., McLaughlin, E. J., Hutchinson, S., and Glover, L. (2019). Escaping the Immune System by DNA Repair and Recombination in African Trypanosomes. *Open Biol.* 9, 190182. doi:10.1098/rsob.190182
- Singh, N., Bergman, Y., Cedar, H., and Chess, A. (2003). Biallelic Germline Transcription at the κ Immunoglobulin Locus. *J. Exp. Med.* 197, 743–750. doi:10.1084/jem.20021392
- Skok, J. A., Brown, K. E., Azuara, V., Caparros, M.-L., Baxter, J., Takacs, K., et al. (2001). Nonequivalent Nuclear Location of Immunoglobulin Alleles in B Lymphocytes. *Nat. Immunol.* 2, 848–854. doi:10.1038/n10901-848
- Soba, P., Zhu, S., Emoto, K., Younger, S., Yang, S.-J., Yu, H.-H., et al. (2007). Drosophila Sensory Neurons Require Dscam for Dendritic Self-Avoidance and Proper Dendritic Field Organization. *Neuron* 54, 403–416. doi:10.1016/j.neuron.2007.03.029
- Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., et al. (2009). Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell* 137, 356–368. doi:10.1016/j.cell.2009.01.052
- Spehr, M., and Munger, S. D. (2009). Olfactory Receptors: G Protein-Coupled Receptors and beyond. *J. Neurochem.* 109, 1570–1583. doi:10.1111/j.1471-4159.2009.06085.x
- Spitz, F. (2019). Chromosomes Come Together to Help Mice Distinguish Odours. *Nature* 565, 439–440. doi:10.1038/d41586-019-00010-6
- Stanne, T. M., and Rudenko, G. (2010). Active VSG Expression Sites in *Trypanosoma Brucei* Are Depleted of Nucleosomes. *Eukaryot. Cel.* 9, 136–147. doi:10.1128/EC.00281-09
- Tan, L., Li, Q., and Xie, X. S. (2015). Olfactory Sensory Neurons Transiently Express Multiple Olfactory Receptors during Development. *Mol. Syst. Biol.* 11, 844. doi:10.15252/msb.20156639
- Tan, L., Xing, D., Daley, N., and Xie, X. S. (2019). Three-dimensional Genome Structures of Single Sensory Neurons in Mouse Visual and Olfactory Systems. *Nat. Struct. Mol. Biol.* 26, 297–307. doi:10.1038/s41594-019-0205-2
- Tas, J. M. J., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J. T., Mano, Y. M., et al. (2016). Visualizing Antibody Affinity Maturation in Germinal Centers. *Science* 351, 1048–1054. doi:10.1126/science.aad3439
- Tasic, B., Nabholz, C. E., Baldwin, K. K., Kim, Y., Rueckert, E. H., Ribich, S. A., et al. (2002). Promoter Choice Determines Splice Site Selection in Protocadherin α and γ Pre-mRNA Splicing. *Mol. Cell.* 10, 21–33. doi:10.1016/S1097-2765(02)00578-6
- Thu, C. A., Chen, W. V., Rubinstein, R., Chevee, M., Wolcott, H. N., Felsovalyi, K. O., et al. (2014). Single-Cell Identity Generated by Combinatorial Homophilic Interactions between α , β , and γ Protocadherins. *Cell* 158, 1045–1059. doi:10.1016/j.cell.2014.07.012
- Turner, C. M. R., and Barry, J. D. (1989). High Frequency of Antigenic Variation in *Trypanosoma Brucei* Rhodensienseinfections. *Parasitology* 99, 67–75. doi:10.1017/S0031182000061035
- Ustaoglu, P., Haussmann, I. U., Liao, H., Torres-Mendez, A., Arnold, R., Irimia, M., et al. (2019). Srrm234, but Not Canonical SR and hnRNP Proteins, Drive Inclusion of Dscam Exon 9 Variable Exons. *RNA* 25, 1353–1365. doi:10.1261/rna.071316.119
- Vassalli, A., Feinstein, P., and Mombaerts, P. (2011). Homeodomain Binding Motifs Modulate the Probability of Odorant Receptor Gene Choice in Transgenic Mice. *Mol. Cell Neurosci.* 46, 381–396. doi:10.1016/j.mcn.2010.11.001
- Vettermann, C., and Schlissel, M. S. (2010). Allelic Exclusion of Immunoglobulin Genes: Models and Mechanisms. *Immunol. Rev.* 237, 22–42. doi:10.1111/j.1600-065X.2010.00935.x
- Victoria, G. D., and Mesin, L. (2014). Clonal and Cellular Dynamics in Germinal Centers. *Curr. Opin. Immunol.* 28, 90–96. doi:10.1016/j.coi.2014.02.010
- Victoria, G. D., Schwickert, T. A., Fooksman, D. R., Kamphorst, A. O., Meyer-Hermann, M., Dustin, M. L., et al. (2010). Germinal Center Dynamics Revealed by Multiphoton Microscopy with a Photoactivatable Fluorescent Reporter. *Cell* 143, 592–605. doi:10.1016/j.cell.2010.10.032
- Von Dannecker, L. E. C., Mercadante, A. F., and Malnic, B. (2006). Ric-8B Promotes Functional Expression of Odorant Receptors. *Proc. Natl. Acad. Sci.* 103, 9310–9314. doi:10.1073/pnas.0600697103
- Wang, J., Ma, X., Yang, J. S., Zheng, X., Zugates, C. T., Lee, C.-H. J., et al. (2004a). Transmembrane/Juxtamembrane Domain-dependent Dscam Distribution and Function during Mushroom Body Neuron Morphogenesis. *Neuron* 43, 663–672. doi:10.1016/j.neuron.2004.06.033
- Wang, J., Zugates, C. T., Liang, I. H., Lee, C.-H. J., and Lee, T. (2002). Drosophila Dscam Is Required for Divergent Segregation of Sister Branches and Suppresses Ectopic Bifurcation of Axons. *Neuron* 33, 559–571. doi:10.1016/s0896-6273(02)00570-6
- Wang, Q.-P., Kawahara, T., and Horn, D. (2010). Histone Deacetylases Play Distinct Roles in Telomeric VSG Expression Site Silencing in African Trypanosomes. *Mol. Microbiol.* 77, 1237–1245. doi:10.1111/j.1365-2958.2010.07284.x

- Wang, S.-Z., Ou, J., Zhu, L. J., and Green, M. R. (2012a). Transcription Factor ATF5 Is Required for Terminal Differentiation and Survival of Olfactory Sensory Neurons. *Proc. Natl. Acad. Sci.* 109, 18589–18594. doi:10.1073/pnas.1210479109
- Wang, S. S., Lewcock, J. W., Feinstein, P., Mombaerts, P., and Reed, R. R. (2004b). Genetic Disruptions of O/E2 and O/E3 Genes Reveal Involvement in Olfactory Receptor Neuron Projection. *Development* 131, 1377–1388. doi:10.1242/dev.01009
- Wang, S. S., Tsai, R. Y. L., and Reed, R. R. (1997). The Characterization of the Olf-1/ebf-like HLH Transcription Factor Family: Implications in Olfactory Gene Regulation and Neuronal Development. *J. Neurosci.* 17, 4149–4158. doi:10.1523/JNEUROSCI.17-11-04149.1997
- Wang, X., Li, G., Yang, Y., Wang, W., Zhang, W., Pan, H., et al. (2012b). An RNA Architectural Locus Control Region Involved in Dscam Mutually Exclusive Splicing. *Nat. Commun.* 3, 1255. doi:10.1038/ncomms2269
- Watson, C. T., Steinberg, K. M., Graves, T. A., Warren, R. L., Malig, M., Schein, J., et al. (2015). Sequencing of the Human IG Light Chain Loci from a Hydattidiform Mole BAC Library Reveals Locus-specific Signatures of Genetic Diversity. *Genes Immun.* 16, 24–34. doi:10.1038/gene.2014.56
- Weiler, E. (1965). Differential Activity of Allelic Gamma-Globulin Genes in Antibody-Producing Cells. *Proc. Natl. Acad. Sci.* 54, 1765–1772. doi:10.1073/pnas.54.6.1765
- Weinberger, A. D., and Weinberger, L. S. (2013). Stochastic Fate Selection in HIV-Infected Patients. *Cell* 155 (3), 497–499. doi:10.1016/j.cell.2013.09.039
- Weiner, J. A., and Jontes, J. D. (2013). Protocadherins, Not Prototypical: a Complex Tale of Their Interactions, Expression, and Functions. *Front. Mol. Neurosci.* 6, 4. doi:10.3389/fnmol.2013.00004
- Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L., and Clemens, J. C. (2004). Alternative Splicing of Drosophila Dscam Generates Axon Guidance Receptors that Exhibit Isoform-specific Homophilic Binding. *Cell* 118, 619–633. doi:10.1016/j.cell.2004.08.021
- Wojtowicz, W. M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S. L. (2007). A Vast Repertoire of Dscam Binding Specificities Arises from Modular Interactions of Variable Ig Domains. *Cell* 130, 1134–1145. doi:10.1016/j.cell.2007.08.026
- Worby, C. A., Simonson-Leff, N., Clemens, J. C., Kruger, R. P., Muda, M., and Dixon, J. E. (2001). The Sorting Nexin, DSH3PX1, Connects the Axonal Guidance Receptor, Dscam, to the Actin Cytoskeleton. *J. Biol. Chem.* 276, 41782–41789. doi:10.1074/jbc.M107080200
- Wu, Q., and Jia, Z. (2021). Wiring the Brain by Clustered Protocadherin Neural Codes. *Neurosci. Bull.* 37, 117–131. doi:10.1007/s12264-020-00578-4
- Wu, Q., and Maniatis, T. (1999). A Striking Organization of a Large Family of Human Neural Cadherin-like Cell Adhesion Genes. *Cell* 97, 779–790. doi:10.1016/S0092-8674(00)80789-8
- Wu, Q., Zhang, T., Cheng, J.-F., Kim, Y., Grimwood, J., Schmutz, J., et al. (2001). Comparative DNA Sequence Analysis of Mouse and Human Protocadherin Gene Clusters. *Genome Res.* 11, 389–404. doi:10.1101/gr.167301
- Xu, B., Shi, Y., Wu, Y., Meng, Y., and Jin, Y. (2019). Role of RNA Secondary Structures in Regulating Dscam Alternative Splicing. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1862, 194381. doi:10.1016/j.bbarm.2019.04.008
- Yamagata, M., and Sanes, J. R. (2008). Dscam and Sidekick Proteins Direct Lamina-specific Synaptic Connections in Vertebrate Retina. *Nature* 451, 465–469. doi:10.1038/nature06469
- Yang, X., Figueiredo, L. M., Espinal, A., Okubo, E., and Li, B. (2009). RAP1 Is Essential for Silencing Telomeric Variant Surface Glycoprotein Genes in Trypanosoma Brucei. *Cell* 137, 99–109. doi:10.1016/j.cell.2009.01.037
- Yang, Y., Zhan, L., Zhang, W., Sun, F., Wang, W., Tian, N., et al. (2011). RNA Secondary Structure in Mutually Exclusive Splicing. *Nat. Struct. Mol. Biol.* 18, 159–168. doi:10.1038/nsmb.1959
- Yang, Z., Huh, S. U., Drennan, J. M., Kathuria, H., Martinez, J. S., Tsuda, H., et al. (2012). Drosophila Vap-33 Is Required for Axonal Localization of Dscam Isoforms. *J. Neurosci.* 32, 17241–17250. doi:10.1523/JNEUROSCI.2834-12.2012
- Yokota, S., Hirayama, T., Hirano, K., Kaneko, R., Toyoda, S., Kawamura, Y., et al. (2011). Identification of the Cluster Control Region for the Protocadherin- β Genes Located beyond the Protocadherin- γ Cluster. *J. Biol. Chem.* 286, 31885–31895. doi:10.1074/jbc.M111.245605
- Yue, Y., Meng, Y., Ma, H., Hou, S., Cao, G., Hong, W., et al. (2016a). A Large Family of Dscam Genes with Tandemly Arrayed 5' Cassettes in Chelicerata. *Nat. Commun.* 7, 11252. doi:10.1038/ncomms11252
- Yue, Y., Yang, Y., Dai, L., Cao, G., Chen, R., Hong, W., et al. (2016b). Long-range RNA Pairings Contribute to Mutually Exclusive Splicing. *RNA* 22, 96–110. doi:10.1261/rna.053314.115
- Zazhytska, M., Kodra, A., Hoagland, D. A., Fullard, J. F., Shayya, H., Omer, A., et al. (20212021). Disruption of Nuclear Architecture as a Cause of COVID-19 Induced Anosmia. *BioRxiv* 1, 430314. doi:10.1101/2021.02.09.430314
- Zhan, X.-L., Clemens, J. C., Neves, G., Hattori, D., Flanagan, J. J., Hummel, T., et al. (2004). Analysis of Dscam Diversity in Regulating Axon Guidance in Drosophila Mushroom Bodies. *Neuron* 43, 673–686. doi:10.1016/j.neuron.2004.07.020
- Zhou, Z., and Belluscio, L. (2012). Coding Odorant Concentration through Activation Timing between the Medial and Lateral Olfactory Bulb. *Cel. Rep.* 2, 1143–1150. doi:10.1016/j.celrep.2012.09.035
- Zhu, H., Hummel, T., Clemens, J. C., Berdnik, D., Zipursky, S. L., and Luo, L. (2006). Dendritic Patterning by Dscam and Synaptic Partner Matching in the Drosophila Antennal Lobe. *Nat. Neurosci.* 9, 349–355. doi:10.1038/nn1652
- Zhuang, Y., Futse, J. E., Brown, W. C., Brayton, K. A., and Palmer, G. H. (2007). Maintenance of Antibody to Pathogen Epitopes Generated by Segmental Gene Conversion Is Highly Dynamic during Long-Term Persistent Infection. *Infect. Immun.* 75, 5185–5190. doi:10.1128/IAI.00913-07
- Zipursky, S. L., and Sanes, J. R. (2010). Chemoaffinity Revisited: Dscams, Protocadherins, and Neural Circuit Assembly. *Cell* 143, 343–353. doi:10.1016/j.cell.2010.10.009
- Zipursky, S. L., Wojtowicz, W. M., and Hattori, D. (2006). Got Diversity? Wiring the Fly Brain with Dscam. *Trends Biochem. Sci.* 31, 581–588. doi:10.1016/j.tibs.2006.08.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Williams, Sikora, Hammer, Amin, Brinjikji, Brumley, Burrows, Carrillo, Cromer, Edwards, Emri, Fergle, Jenkins, Kaushik, Maydan, Woodard and Clowney. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.