Measuring Causal Specificity

Supplementary Online Materials

Paul E. Griffiths[1], Arnaud Pocheville[1], Brett Calcott[2], Karola Stotz[3], Hyunju Kim[4] and Rob Knight [5]

[1]Department of Philosophy and Charles Perkins Centre, University of Sydney, NSW 2006, Australia

[2]ASU/SFI Center for Complex Biosocial Systems School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

[3]Department of Philosophy, Macquarie University, NSW 2109, Australia

[4]Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ 85287, USA

[5]Howard Hughes Medical Institute, and Departments of Chemistry & Biochemistry and Computer Science, and BioFrontiers Institute, University of Colorado at Boulder, Boulder, CO 80309, USA

July 30, 2015

Philosophy of Science 82, no. 4 (in press).

1 The effect of transcription probability

Here we derive the equations of the curves in Figure 7 (reproduced below as Figure 2) describing the effect of transcription probability on several informational measures on RNA, DNA and transcription. For the ease of presentation, we will ignore splicing.

To ease reading, we will write the variables RNA as R (with values r_i), transcription as T (with values t_h), and DNA as D (with values d_j). Again, hats on variables mean that their values are fixed by a surgical intervention.

1.1 The mutual information between RNA and transcription

We suppose that if there is no transcription ($h = 0$), there is no RNA strand produced ($i = 0$), while if there is transcription ($h = 1$), there is one RNA strand produced among n possible variants ($i = 1 \dots n$). This implies that once a given value for RNA is obtained (either $i = 0$, i.e. absence, or $i = 1 \dots n$) we also know whether transcription was *on* or *off*. In other words, the joint probability for RNA and transcription is given as follows (see Figure 1):

$$p(r_i, \widehat{t_h}) = \begin{cases} p(r_0), & \text{if } h = 0 \text{ and } i = 0. \\ p(r_i), & \text{for } h = 1 \text{ and } i = 1, 2, \dots, n. \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Also, by computing the marginal probability of transcription, $p(\widehat{t_h}) = \sum_{i=0}^{n} p(r_i, \widehat{t_h})$, we can obtain that $p(\widehat{t_0}) = p(r_0, \widehat{t_0})$ and $p(\widehat{t_1}) = \sum_{i=1}^{n} p(r_i, \widehat{t_1})$. Therefore,

$$p(\widehat{t_0}) = p(r_0) \quad \text{and} \quad p(\widehat{t_1}) = \sum_{i=1}^{n} p(r_i) \tag{2}$$

Now, using (1) and (2), we can compute the mutual information between

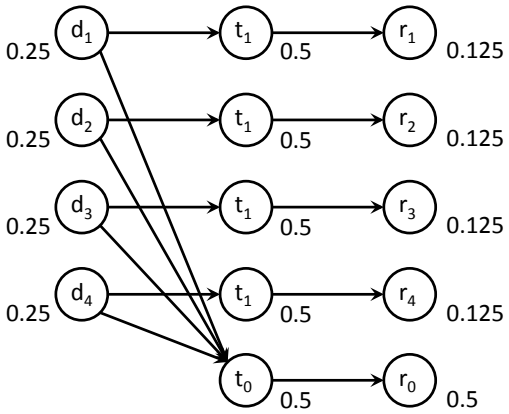

Figure 1: Diagram showing events with non-null probabilities in our model of transcription, when splicing is ignored. Transcription can be either *on* ($h = 1$), in which case a DNA strand j will deterministically lead to a RNA strand j, or *off* ($h = 0$), in which case any DNA strand will lead to a null RNA. (Probabilities assigned to events are for illustratory purpose only, but notice that $p(t_0)$ and $p(t_1)$ sum to 1.)

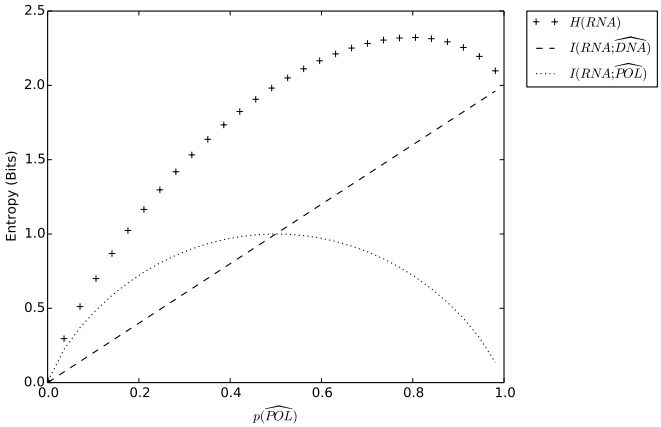

Figure 2: Effects of changing probability of transcription on several informational measures: the entropy of RNA (the effect), the mutual information between RNA and DNA, and the mutual information between RNA and the presence of polymerase.

RNA and transcription.

$$
\begin{aligned}
I\left(R;\widehat{T}\right) &= \sum_{h=0}^{1}\sum_{i=0}^{n} p\left(r_i,\widehat{t_h}\right)\log\frac{p\left(r_i,\widehat{t_h}\right)}{p\left(r_i\right)p\left(\widehat{t_h}\right)} && (3)\\
&= p\left(r_0,\widehat{t_0}\right)\log\frac{p\left(r_0,\widehat{t_0}\right)}{p\left(r_0\right)p\left(\widehat{t_0}\right)} + \sum_{i=1}^{n} p\left(r_i,\widehat{t_1}\right)\log\frac{p\left(r_i,\widehat{t_1}\right)}{p\left(r_i\right)p\left(\widehat{t_1}\right)} && (4)\\
&= p\left(r_0\right)\log\frac{p\left(r_0\right)}{p\left(r_0\right)p\left(\widehat{t_0}\right)} + \sum_{i=1}^{n} p\left(r_i\right)\log\frac{p\left(r_i\right)}{p\left(r_i\right)p\left(\widehat{t_1}\right)} && (5)\\
&= p\left(r_0\right)\log\frac{1}{p\left(\widehat{t_0}\right)} + \sum_{i=1}^{n} p\left(r_i\right)\log\frac{1}{p\left(\widehat{t_1}\right)} && (6)\\
&= p\left(r_0\right)\log\frac{1}{p\left(\widehat{t_0}\right)} + \left(\sum_{i=1}^{n} p\left(r_i\right)\right)\log\frac{1}{p\left(\widehat{t_1}\right)} && (7)\\
&= p\left(\widehat{t_0}\right)\log\frac{1}{p\left(\widehat{t_0}\right)} + p\left(\widehat{t_1}\right)\log\frac{1}{p\left(\widehat{t_1}\right)} && (8)\\
&= H\left(\widehat{T}\right) && (9)
\end{aligned}
$$

That $I\left(R;\widehat{T}\right) = H\left(\widehat{T}\right)$ simply reflects that there is a bijection between having transcription set to *on* (respectively *off*) and obtaining some *non-null* (respectively *null*) RNA. In other words, none of the values for transcription lead to convergent results: there is no loss of information about transcription when it occurs (or not).

1.2 The mutual information between RNA and DNA

We suppose that if there is transcription ($h = 1$), a given strand of DNA ($j = 1...n$) will deterministically lead to a given strand of RNA ($i = 1...n$). If there is no transcription ($h = 0$), any strand of DNA will lead to no RNA ($i = 0$) (see Figure 3). In other terms, there is a bijection between DNA and RNA if and only if transcription is on, otherwise all values of DNA lead to the same null result. We also suppose that state of the polymerase and the choice of a DNA strand to transcribe are independent events.

We begin with:

$$I\left(R;\widehat{D}\right)=\sum_{i=0}^{n}\sum_{j=1}^{n}p\left(r_i,\widehat{d}_j\right)\log\frac{p\left(r_i,\widehat{d}_j\right)}{p\left(r_i\right)p\left(\widehat{d}_j\right)} \tag{10}$$

We will consider now how this measure behaves when we take into account the probability of transcription.

To simplify writing, we will first notice that many joint events have null probabilities, which makes them cancel out in the calculus of mutual information. These joint events are $(r_{i>0}, d_{j\neq i})$: it is impossible to get another strand of RNA than the one the DNA strand codes for (whatever the transcription state, see Figure 1).

Thus, without loss of generality, we can write, splitting the cases with *non-null* ($i>0$) and *null* ($i=0$) RNA:

$$I\left(R;\widehat{D}\right)=\sum_{i=1}^{n}p\left(r_i,\widehat{d}_i\right)\log\frac{p\left(r_i,\widehat{d}_i\right)}{p\left(r_i\right)p\left(\widehat{d}_i\right)}+\sum_{j=1}^{n}p\left(r_0,\widehat{d}_j\right)\log\frac{p\left(r_0,\widehat{d}_j\right)}{p\left(r_0\right)p\left(\widehat{d}_j\right)} \tag{11}$$

Using the diagram in Figure 1, we can easily see the following relationships:

(a) $p(\widehat{d}_i|r_i)=1,$ if $i>0.$

(b) $p(\widehat{d}_j|r_0)=p(\widehat{d}_j),$ for $j=1,\ldots n.$

(c) $p(r_i|\widehat{d}_i)=p(\widehat{t}_1),$ if $i>0.$

Using these relationships, we can simplify $I(R;\widehat{D})$ as follows:

$$\begin{aligned} I\left(R;\widehat{D}\right) &= \sum_{i=1}^{n} p\left(r_i,\widehat{d}_i\right)\log\frac{p\left(r_i,\widehat{d}_i\right)}{p\left(r_i\right)p\left(\widehat{d}_i\right)} \\ &\quad + \sum_{j=1}^{n} p\left(r_0,\widehat{d}_j\right)\log\frac{p\left(r_0,\widehat{d}_j\right)}{p\left(r_0\right)p\left(\widehat{d}_j\right)} \qquad (12) \\ &= \sum_{i=1}^{n} p\left(r_i,\widehat{d}_i\right)\log\frac{p\left(\widehat{d}_i|r_i\right)}{p\left(\widehat{d}_i\right)} \\ &\quad + \sum_{j=1}^{n} p\left(r_0,\widehat{d}_j\right)\log\frac{p\left(\widehat{d}_j|r_0\right)}{p\left(\widehat{d}_j\right)} \qquad (13) \end{aligned}$$

Due to relationships (a) and (b),

$$\begin{aligned} I\left(R;\widehat{D}\right) &= \sum_{i=1}^{n} p\left(r_i,\widehat{d}_i\right)\log\frac{1}{p\left(\widehat{d}_i\right)} + \sum_{j=1}^{n} p\left(r_0,\widehat{d}_j\right)\log\frac{p\left(\widehat{d}_j\right)}{p\left(\widehat{d}_j\right)} \qquad (14) \\ &= \sum_{i=1}^{n} p\left(r_i,\widehat{d}_i\right)\log\frac{1}{p\left(\widehat{d}_i\right)} \qquad (15) \\ &= \sum_{i=1}^{n} p\left(r_i|\widehat{d}_i\right)p\left(\widehat{d}_i\right)\log\frac{1}{p\left(\widehat{d}_i\right)} \qquad (16) \end{aligned}$$

Due to relationship (c),

$$\begin{aligned} I\left(R;\widehat{D}\right) &= \sum_{i=1}^{n} p(\widehat{t_1})p\left(\widehat{d}_i\right)\log\frac{1}{p\left(\widehat{d}_i\right)} \qquad (17) \\ &= p(\widehat{t_1})H(\widehat{D}) \qquad (18) \end{aligned}$$

This equation reflects the fact that the informativity of DNA is conditional upon the presence of transcription. If transcription were *always* on, there would be a bijection between DNA and RNA. However, when the transcription is sometimes off, there is a loss of information between DNA and the RNA outputs, as several strands of DNA can lead to the same result (no RNA) when there is no transcription. The information loss is simply this

part of DNA entropy which is not present in the mutual information between DNA and RNA, that is, $H\left(\widehat{D}\,|R\right)$:

$$H\left(\widehat{D}\,|R\right) = H\left(\widehat{D}\right) - I\left(R;\widehat{D}\right) \quad (19)$$

$$= \left(1-p\left(\widehat{t_1}\right)\right) H\left(\widehat{D}\right) \quad (20)$$

1.3 The entropy of RNA

Here we derive the entropy of RNA in terms of mutual information between RNA and DNA and the entropy of transcription. We again split between the cases where there is transcription ($\widehat{t_1}$) or none ($\widehat{t_0}$). We again use the fact that and that $p\left(r_i\right) = p\left(\widehat{d_i}\right) p\left(\widehat{t_i}\right)$. We also remark that $\sum_{i=1}^{n} p\left(\widehat{d_i}\right) p\left(\widehat{t_1}\right)$ sums to $p\left(\widehat{t_1}\right)$.

$$H\left(R\right) = -\sum_{i=0}^{n} p\left(r_i\right) \log p\left(r_i\right) \quad (21)$$

$$= -\sum_{i=1}^{n} p\left(\widehat{d_j}\right) p\left(\widehat{t_1}\right) \log p\left(\widehat{d_j}\right) p\left(\widehat{t_1}\right) - p\left(r_0\right) \log p\left(r_0\right) \quad (22)$$

$$= -p\left(t_1\right) \left(\sum_{i=1}^{n} p\left(\widehat{d_j}\right) \log p\left(\widehat{d_j}\right) + \sum_{i=1}^{n} p\left(\widehat{d_j}\right) \log p\left(\widehat{t_1}\right) \right) - p\left(r_0\right) \log p\left(r_0\right) \quad (23)$$

$$= -p\left(t_1\right) \sum_{i=1}^{n} p\left(\widehat{d_j}\right) \log p\left(\widehat{d_j}\right) - p\left(\widehat{t_1}\right) \log p\left(\widehat{t_1}\right) - p\left(\widehat{t_0}\right) \log p\left(\widehat{t_0}\right) \quad (24)$$

We recognize:

$$H\left(R\right) = p\left(\widehat{t_1}\right) H\left(\widehat{D}\right) + H\left(\widehat{T}\right) \quad (25)$$

$$= I\left(R;\widehat{D}\right) + I\left(R;\widehat{T}\right) \quad (26)$$

2 The mutual information between RNA and splicing

2.1 When transcription is always on

Here we derive the equations for the mutual information between RNA and splicing.

For the sake of simplicity, we shall first ignore transcription probability and assume that $p\left(\widehat{t_1}\right) = 1$. This amounts to relaxing the conditionalisation upon transcription.

In the model considered here the splicing factor variants are recruited only once a given strand of DNA has been transcribed. In addition, we suppose that the transcription of a given DNA strand opens a set of possibilities among a *proper* set of splicing factors (see Figure 3). This entails that the information in splicing $H\left(\widehat{S}\right)$ contains all the information in DNA, $H\left(\widehat{D}\right)$:

$$H\left(\widehat{D}, \widehat{S}\right) = H\left(\widehat{S}\right) \tag{27}$$

In addition, we consider a bijective relationship between splicing factors and RNA variants. This bijection entails that the mutual information between RNA and splicing is equal to the self-information of splicing (that is, the entropy of splicing). We can then decompose the entropy of splicing according to well known chain rules:

$$\begin{aligned} I\left(R; \widehat{S}\right) &= I\left(\widehat{S}; \widehat{S}\right) && (28)\\ &= H\left(\widehat{S}\right) && (29)\\ &= H\left(\widehat{D}, \widehat{S}\right) && (30)\\ &= H\left(\widehat{S} \middle| \widehat{D}\right) + H\left(\widehat{D}\right) && (31) \end{aligned}$$

From equation (18), we know that $H\left(\widehat{D}\right) = I\left(R; \widehat{D}\right)$, assuming that transcription always occurs. In addition, the bijection between splicing and RNA (including the null value) entails that the conditional entropy of splicing

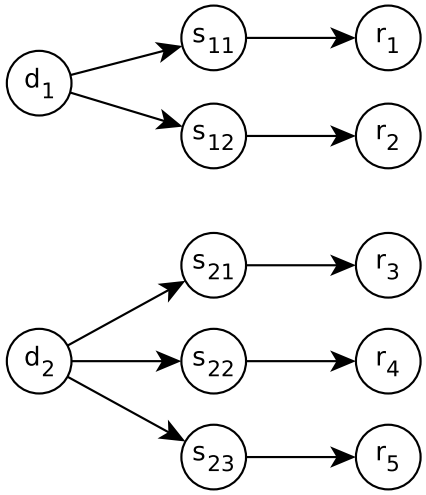

Figure 3: Diagram of our model of splicing, when transcription is assumed to be on. A DNA strand deterministically leads to a proper set of splicing factor variants, each of them deterministically leading to a proper RNA strand.

(conditioned on DNA) is the conditional mutual information of splicing and RNA: $H\left(\widehat{S}\middle|\widehat{D}\right) = I\left(R;\widehat{S}\middle|\widehat{D}\right)$ (as an immediate calculation would show). We thus can rewrite equation (31) as:

$$I\left(R;\widehat{S}\right) = I\left(R;\widehat{S}\middle|\widehat{D}\right) + I\left(R;\widehat{D}\right) \tag{32}$$

Readers familiar with information theory will recognize the decomposition of the mutual information $I\left(R;\widehat{S},\widehat{D}\right)$ which happens to be, in this particular example, equal to $I\left(R;\widehat{S}\right)$. That is, knowing the value of DNA does not bring us any information as regards RNA in addition to knowing the value of splicing. Notice equation (32) also provides a decomposition of the entropy of splicing, that is, $H(S) = I\left(R;\widehat{S}\right)$ in virtue of the bijection between RNA and splicing.

2.2 When transcription can be either on or off

For the sake of completeness, we now give equation (32) in a version taking into account the probability of transcription. The reasoning is grounded on the hypothesis that a given splicing factor occurs only when there is transcription and a given DNA strand has been choosen. Then, decomposition of $I(R;\widehat{S})$ gives:

$$\begin{aligned} I\left(R;\widehat{S}\right) &= I\left(\widehat{S};\widehat{S}\right) && (33)\\ &= H\left(\widehat{S}\right) && (34)\\ &= H\left(\widehat{S}\left|\widehat{D},\widehat{T}\right.\right) + I\left(\widehat{S};\widehat{D},\widehat{T}\right) && (35)\\ &= H\left(\widehat{S}\left|\widehat{D},\widehat{T}\right.\right) + I\left(\widehat{S};\widehat{D}\left|\widehat{T}\right.\right) + I\left(\widehat{S};\widehat{T}\right) && (36) \end{aligned}$$

Again, we take advantage of the bijection between splicing and RNA, to replace $H\left(\widehat{S}\left|\widehat{D},\widehat{T}\right.\right) = I\left(R;\widehat{S}\left|\widehat{D},\widehat{T}\right.\right)$. We also take advantage of the fact that there is no interaction information between DNA, RNA, and transcription, that is, $I\left(R;\widehat{D}\left|\widehat{T}\right.\right) = I\left(R;\widehat{D}\right)$. This can be shown with the calculation sketched below. We again use relationship (1) to simplify, hence if $i > 0$ we have $p(r_i, d_j, t_1) = p(r_i, d_j)$ and $p(r_i, t_1) = p(r_i)$. A similar replacement method would hold for $i = 0$, but we directly simplify this term

as it is null.

$$
\begin{aligned}
I\left(R;\widehat{D}\left|\widehat{T}\right.\right) &= \sum_{h=0}^{1} p(\widehat{t_h}) \sum_{j=0}^{n}\sum_{i=1}^{m} p\left(r_i,\widehat{d_j}\left|\widehat{t_h}\right.\right) \log \frac{p\left(r_i,\widehat{d_j}\left|\widehat{t_h}\right.\right)}{p\left(r_i\left|\widehat{t_h}\right.\right) p\left(\widehat{d_j}\left|\widehat{t_h}\right.\right)} && (37)\\
&= p(\widehat{t_0}) \sum_{j=0}^{n} p\left(r_0,\widehat{d_j}\left|\widehat{t_0}\right.\right) \log \frac{p\left(r_0,\widehat{d_j}\left|\widehat{t_0}\right.\right)}{p\left(r_0\left|\widehat{t_0}\right.\right) p\left(\widehat{d_j}\left|\widehat{t_0}\right.\right)} && (38)\\
&\quad + p(\widehat{t_1}) \sum_{j=0}^{n}\sum_{i=1}^{m} p\left(r_i,\widehat{d_j}\left|\widehat{t_1}\right.\right) \log \frac{p\left(r_i,\widehat{d_j}\left|\widehat{t_1}\right.\right)}{p\left(r_i\left|\widehat{t_1}\right.\right) p\left(\widehat{d_j}\left|\widehat{t_1}\right.\right)} && (39)\\
&= 0 + \sum_{j=0}^{n}\sum_{i=1}^{m} p\left(r_i,\widehat{d_j},\widehat{t_1}\right) \log \frac{p\left(r_i,\widehat{d_j},\widehat{t_1}\right)}{p(\widehat{t_1}) p\left(r_i\left|\widehat{t_1}\right.\right) p\left(\widehat{d_j}\right)} && (40)\\
&= \sum_{j=0}^{n}\sum_{i=1}^{m} p\left(r_i,\widehat{d_j}\right) \log \frac{p\left(r_i,\widehat{d_j}\right)}{p\left(r_i\right) p\left(\widehat{d_j}\right)} && (41)\\
&= I\left(R;\widehat{D}\right) && (42)
\end{aligned}
$$

Injecting these terms in equation (36), we obtain:

$$
\begin{aligned}
I\left(R;\widehat{S}\right) &= H\left(\widehat{S}\left|\widehat{D},\widehat{T}\right.\right) + I\left(\widehat{S};\widehat{D}\right) + I\left(\widehat{S};\widehat{T}\right) && (43)\\
&= H\left(\widehat{S}\left|\widehat{D},\widehat{T}\right.\right) + p\left(\widehat{t_1}\right) H\left(\widehat{D}\right) + H\left(\widehat{T}\right) && (44)
\end{aligned}
$$

Again, noticing that $H\left(\widehat{S}\left|\widehat{D},\widehat{T}\right.\right) = I\left(R;\widehat{S}\left|\widehat{D},\widehat{T}\right.\right)$, we retrieve an equation similar to equation (32):

$$
I\left(R;\widehat{S}\right) = I\left(R;\widehat{S}\left|\widehat{D},\widehat{T}\right.\right) + I\left(R;\widehat{D}\right) + I\left(R;\widehat{T}\right) \tag{45}
$$

Readers familiar with information theory will recognize the decomposition of the mutual information $I\left(R;\widehat{S},\widehat{D},\widehat{T}\right)$ which happens to be, in this particular example, equal to $I\left(R;\widehat{S}\right)$. That is, knowing the value of DNA and transcription does not bring us any more information as regards RNA than just knowing the value of splicing. Notice that similarly to equation (32) in the case where transcription is always on, equation (45) provides a decomposition of the entropy of splicing, that is, $H(S) = I\left(R;\widehat{S}\right)$ in virtue of the bijection between RNA and splicing.

To wrap up, in this model transcription adds variation in the set of splicing factor variants (the absence of any factor now belongs to the set of possibilities), which is independent from DNA.

3 Alternative splicing in Drosophila Dscam

The Drosophila receptor DSCAM (Down Syndrome Cell Adhesion Molecule), a member of the immunoglobulin (Ig) superfamily, is a remarkable example of homophilic binding specificity that functions in important biological processes, such as innate immunity and neural wiring. In insects and also crustaceans (e.g. Daphnia) 4 of the 24 exons of the *Dscam* gene are arranged in large tandem arrays, whose regulation is an example of mutually exclusive splicing. In Drosophila one block has 2 exons - leading to 1 of 2 alternative transmembrane segments, the others contain respectively 12, 48 and 33 alternative exons - leading to 19,008 different ecto-domains. Together they produce, 38,016 alternative protein isoforms, within a genome of 15,016 protein-coding genes [1]. There are several interesting aspects about this case:

1. For each block of exons there seem to exist a unique mechanism that ensures that exclusively only one of the alternative axons is included in the final transcript. Only two of the mechanisms are known in some detail. Researchers have identified specific cis-acting sequences and trans-acting splicing factors that tightly regulate splicing of exon 4.2, but for most others the details are again not yet known [4, 3].

2. It is not only the large number of alternative transcripts that allow for high diversity of functions, but in addition most alternative exons are expressed in neurons and found in many combinations. Neurons express up 50 variants at a time, which makes for an even larger combinatorial spectrum of neuron differentiation. This ensures that branches

from different neurons will share, at most, a few isoforms in common. This diversity of function enables branches of neurons to distinguish between sister branches and branches of other neurons, and also for patterning of neural circuits [8].

3. There seem to be distinct ways of regulating isoforms in the two different functions. For self-recognition purposes, neurons seem to express DSCAM isoforms in a stochastic yet biased fashion. Which isoform is expressed in a single neuron is unimportant as long as it sufficiently different from its neighbour. It might simply be an indirect consequence of the expression of different splicing factors in different neurons that leads to this bias. For appropriate branching patterns, however, the research to date suggests that the expression of *Dscam* isoforms in some neurons is under tight developmental control. So we find a controlled mix of stochasticity and regulation in the expression of *Dscam* in drosophila [9].

4. *Dscam* is homologous between almost all animals, which places its origin to over 600 million years ago before the split between the deuterostomes and protostomes [2]. But while in vertebrates their two homologous genes, *Dscam* and *Dscam*L1 do not encode multiple isoforms, in arthropods the single gene is highly enriched with alternative exons. That leads to the interesting hypothesis that while in simple animals cell adhesion and cell recognition is controlled by complex genes, in complex animals this is done by relatively simple genes. This raises the question of how to address the difficulty of accounting for a molecular diversity large enough to provide specificity for the extraordinary large number of neurons in the more complex vertebrate brains [5].

Vertebrates seem to manage their increase in cell recognition specificities through the combinatorial association of different recognition systems

such as gene duplication and the successive divergence of other loci, and via the graded expression of recognition proteins [9]. There exists a large range of cell adhesion, recognition and surface receptor genes in vertebrates: the calcium-independent Ig superfamily, and calcium-dependent integrins, cadherins, and selectins. The human immunoglobulins (Ig) are the products of three unlinked sets of genes: the immunoglobulin heavy (IGH), the immunoglobulin (IGK), and the immunoglobulin (IGL) genes, with a total of about 150 functional genes. A large number of cadherin superfamily genes have been identified to date, and most of them seem to be expressed in the CNS. At least 80 members of the cadherin superfamily have been shown to be expressed within a single mammalian species. Integrins have two different chains, the (alpha) and (beta) subunits of which mammals possess eighteen and eight subunits, while Drosophila has five and two subunits.

This genetic diversity is combined with complex regulatory patterns. One example are the neurexin and neuroligin proteins in humans which are all encoded by multiple genes. Neurexin is encoded by three genes controlled each by two promoters which produce 6 main forms of neurexin. Both genes display relative extensive alternative splicing, a process that can potentially generate thousands of neurexin isoforms alone [2, 6]. Splice form diversity is most extensive in the mammalian brain [7].

4 References

1. Alicia M. Celottoa and Brenton R. Graveley. Alternative splicing of the drosophila dscam pre mRNA is both temporally and spatially regulated. Genetics, 159:599–608, 2001.

2. Mack E. Crayton III, Bradford C. Powell, Todd J. Vision, and Morgan C. Giddings. Tracking the evolution of alternatively spliced exons within the dscam family. BMC Evolutionary Biology, 6(16), 2006.

3. Brenton R. Graveley. Mutually exclusive splicing of the insect dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell, 123:65–73, 2005.

4. Jung Woo Park and Brenton R. Graveley. Complex alternative splicing. In Benjamin J. Blencowe and Brenton R. Graveley, editors, Alternative Splicing in the Postgenomic Era, number 623 in Advances in Experiemental Medicine and Biology, pages 50–63. Landes Bioscience and Springer Science+Business Media, 2007.

5. Dietmar Schmucker and Brian Chen. Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. Genes Dev., 23:147–156, 2009.

6. Barbara Treutleina, Ozgun Gokcec, Stephen R. Quakea, and Thomas C. Sudhof. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. PNAS, 11(13):E1291–E1299, 2014.

7. Julian P. Venables, Jamal Tazi, and Franois Juge. Regulated functional alternative splicing in drosophila. Nucleic Acids Research, 40(1):1–10, 2012.

8. Woj M. Wojtowicz, John J. Flanagan, S. Sean Millard, S. Lawrence Zipursky, and James C. Clemens. Alternative splicing of drosophila dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell, 118(5):619–633, 2004.

9. S. Lawrence Zipursky, Woj M. Wojtowicz, and Daisuke Hattori. Got diversity? wiring the fly brain with dscam. Trends in Biochemical Sciences, 31(10):581–588, 2006.