AMJ Australasian Medical Journal [AMJ 2017;10(8):701-715]

A new view of genetic mutations

Jean-François Picimbon

QILU University of Technology, School of Bioengineering, Jinan 250353, China

REVIEW

Please cite this paper as: Picimbon JF. A new view of genetic mutations. AMJ 2017;10(8):701–715.
https://doi.org/10.21767/AMJ.2017.3096

Corresponding Author:
Jean-François Picimbon
QILU University of Technology, School of Bioengineering
Jinan 250353, China
Email: jfpicimbon@gmail.com

ABSTRACT

I discussed RNA editing mechanisms leading to the diversity of protein; new mechanisms escape from common general knowledge in biology that one gene encoding for one single protein. I also gave two known examples of RNA editing mechanisms that generate a high level of protein diversity: chemosensory proteins (CSPs) and odour binding protein (OBP) families. It is worthy to note that these RNA variants are not disease-causing mutations but rather an evolutionary mechanism in microorganisms and insects.

Furthermore, these genetic variants presenting at the RNA level are not lethal and tissue-specific. One example, the sex pheromone (bombykol) gland of the silkworm moth *Bombyx mori*, is the preferential site of RNA mutations. Moreover, I have perhaps demonstrated that RNA splicing, editing and/or protein recombination are the crucial mechanisms for the totipotent cells to differentiate into a specific cell type in a given tissue. I give a global view of the basic phenomenon of eukaryotic cells and present a new perspective for the treatment of genetic diseases.

Key Words
Chemosensory protein, odorant-binding protein, RNA editing, pheromone gland, cell potency

Implications for Practice:

1. What is known about this subject?

Mutations are limited to A-to-I and/or C-to-U conversion. Changing one amino acid in the protein structure may lead to lethal conditions.

2. What new information is offered in this report?

The high diversity of CSP-RNA and isoform variants suggests the occurrence of mechanisms other than A-to-I and C-to-U conversion to underlie protein diversity. Changing one or a few amino acids in the protein structure may not be so fatal, but may introduce new functions in the protein family, a prelude to multi-potency.

3. What are the implications for research, policy, or practice?

Research should be made for on-tissue RNA/protein mutation and new function, particularly in multi-potent cells. A cell, a tissue or an organism that can tolerate such a load or panoply of mutations and from it produce huge protein heterogeneity, yet adapts to new environment and sustains development as well as natural evolution. The mechanisms behind these mutations yet to be found might offer new strategies to rescue damaged tissues by means of change-oriented self-renewal capacity and multi-potency in stem cells for future practice of regenerative medicine.

Background

One gene coding for one single protein is textbook, a common general knowledge in biology spread from generations to generations of new students and scientists. This comes from the RNA structure that leaves the nucleus and reaches the ribosome, where it is processed to a protein structure.[1] One gene-one protein is a misconception of post-transcriptional events (epigenetics) after the discovery of alternative or differential splicing in bacteria.[2,3] Therefore, cutting off the RNA to produce different proteins might be extremely ancient, emerging from a prebiotic time or an earlier evolution of microorganisms. Even though the diversity of spliced variants from a single gene may be rather limited by the size of the gene, it earlier rejected the dogma and proposed an evolutionary model in which a cell

or tissue increased protein diversity in a remarkable manner by reorganizing functional modules without significantly increasing genome size.[4]

The answer for an evolutionary mechanism that conveys RNA and protein diversity from a single gene is not only alternative gene splicing. It needs to also refer to a phenomenon of pinpointed replacement of nucleotide base on an RNA strand, i.e., RNA editing, which can even more significantly expand the protein repertoire independently of the gene structure.[5] RNA editing is currently known to be limited to alteration of very specific sites in the RNA sequence. The adenosine deaminase acting on an RNA (ADAR) protein family only converts adenosine to inosine in double-stranded RNA duplexes.[6-8] Another catalytic component of RNA editing includes activation-induced cytidine deaminase (AID) and apobec-1 enzyme, a mammalian cytidine deaminase that can act on DNA and RNA single strands.[9,10] Both pyrimidine and purine conversion are known to be tissue-dependent, mainly expressing itself during forebrain development and in the intestine or the liver of mammals including humans.[11-13]

RNA editing in insect binding protein families

Changes in gene expression due to RNA editing also exists in insects as described in the fruit fly *Drosophila melanogaster* and the German cockroach sodium channels.[14] In addition to ion channel receptors and genes involved in neuronal excitability, RNA editing is necessary for a much larger number of protein gene families including genes involved in ion homeostasis, signal transduction, KP transposable elements, so it is not only limited to genes involved in neuronal excitability.[14-18]

Our work on various binding protein families in the silkworm moth *Bombyx mori* brings the notion of RNA editing throughout the whole insect body, and particularly in the cells of the sex pheromone gland, which synthesizes the sex-attractant, Bombykol (Figure 1).[19-23] A high level of RNA mutations and correlatively a high number of peptide variants occur for the chemosensory protein (CSP) gene family in the female sex pheromone gland of the silkworm.[19-23]

Chemosensory proteins or CSPs are proteins believed to play an important role in lipid metabolism and thereby pheromone biosynthesis. Some CSPs have been shown to bind long fatty acid lipid chains, i.e., linoleic acid, which might serve not only in pheromone biosynthetic pathways, but also developmental regulation and immune response.[24] Other CSPs have been shown to bind completely different chemical structures, a type of phenolic compound with conjugated system (cinnamaldehyde),[24] emphasizing binding capacity for an extremely broad range of ligands in many various compartments of the insect body.[25-27] CSPs probably take part in the assembly of various multi-molecular protein complexes, building the machinery necessary for the cell to retain multiple essential crucial functions. These multi-functional roles of CSPs are applied not only to insect cells, but also to prokaryotic cells of bacterial microorganisms.[28]

Such diversity of CSPs functions might be largely due to RNA editing and protein mutation as demonstrated using the silkworm as a model study.[19,20,29] The experiments were done because we observed nucleotide substitutions in CSP clones and numerous N-terminal sequences for the same protein. This was done sequencing pools of individuals on the way to characterize the CSPs in cockroaches, locusts and moths.[30-33] Later, we therefore set out to look out at the nucleotide substitution and the protein diversity in insect CSP genes at the level of individuals.[19] For each individual, we compared ten genomic DNA clones and ten cDNA clones from the antennal, leg, head, wing and pheromone gland tissue for each gene investigated.[19] Finally, we checked for the expression of CSP variant isoforms using immunoblots and peptide sequencing.[19] The main results are in *Bombyx* a surprising amount of DNA/RNA mismatches (or mutations) that could not just be the result of reverse transcriptase, PCR or sequencing errors. Analysing four CSP genes (*BmorCSP1*, *BmorCSP2*, *BmorCSP4* and *BmorCSP14*) in several tissues at the individual level, we found stop codon, deletion, insertion, high level of conventional A-to-I and C-to-U as well as U-to-A, U-to-G and C-to-I mutations expressed in a tissue-specific manner on the RNA. The female sex pheromone gland in particular, was a crucial main site for prominently high diverse RNA editing levels (Figure 2).[19] No mutations were found analysing the cloned genomic DNA loci of these genes. Furthermore, for most *in vitro* cDNA/RNA amplification, the number of incorporated errors during reverse transcription was likely negligible for one main reason: the reverse transcription process does not amplify introduced errors in the cDNA strand.[19] Also, analysing genomic DNA and cDNA clones of other genes such as *actin* in various tissues did not lead to such incredible amount of mismatches on the RNA clone, strongly suggesting that most of these cloned mutations are not *in vitro* artifacts, but exist *in vivo* as subtle nucleotide base replacements for a very peculiar type of cell, differentially expressed and/or subjected to a very peculiar pattern of regulation, generating multiple RNA variants from a single gene and thereby a panoply of new functional

protein isoforms.[19] A similarly large and diverse spectrum of mutations was found in *odour-binding proteins* (*OBPs*) including *pheromone-binding protein* (*PBP*), *general odorant-binding protein* (*GOBP*), *PBP-related protein*, *sericotropin* and *protein B1*.[19,28,29,34,35] This common phenomenon as highlighted by the finding of mutations in *CSPs* but also in *OBPs* suggests that RNA editing at multiple sites causes more than A-to-I or C-to-U changes in order to increase the repertoire of multi-protein transporters, binding proteins and/or molecular carriers.

Protein editing in insect binding protein family

Our detailed analysis of mutations in insect binding proteins brings the notion of protein change not only at the level of RNA editing, but also later after the polypeptide chain elongation (Figure 2). This comes from the finding of subtle pinpointed single point mutations at the RNA level, but of peptide fragments with specific amino acid and/or amino acid motif at the protein level. Frameshift mutations leading to truncated proteins but no triplet codons insertion are found for *CSPs*.[19] Sequences of CSP peptide fragments from libraries specific to the female sex pheromone gland reveal changes in amino acids, that includes inversions, deletions and even a complete replacement of the N-terminal α-helical arm (α1), which could not be explained by the activity of ADAR, ADI or apobec enzymes.[19] Even if such editing enzymes are involved, their accuracies cannot be high enough to induce such a drastic but precise change in the functional protein structure.[19,20,29] Also, very intriguingly, each of these drastic mutations in the N-terminus did not result from a unique combination or re-arrangement of amino acids, but they resulted from a variety of combinations of different amino acid residues added in the motif.[19,20] This would suggest a cellular mechanism capable of producing an enormous repertoire of protein variants, not from a single RNA strand but from a single protein stretch of 10-20 amino acid residues, which are crucial for the protein function. Interestingly, there was a load of supplementary cysteines (Cys) added by protein mutation specifically in the N-terminal arm (α1) of the CSP structure.[19,20] In addition, there were a large number of CSP fragments characterized by insertion of glycine residue (+Gly) in the flank of pre-existing Cys at specific locations in the protein structure (Figure 2).[19,20] +Gly enrolled Cysteine at position 29, 36 and/or 55, but never at Cysteine 58 (Figure 2).[19,20]

These types of mutations, i.e., protein mutations, particularly Gly insertions, were never observed previously. This is the first time it was described. The peptide fragments from a protein band specific to the pheromone gland (where the degree of mutation was remarkably high) were sequenced by nano-liquid chromatography coupled to tandem mass spectrometry (LC/MS/MS).[19] The protein band did not correspond to the mature proteins but rather to a truncated isoform. The band showed expression of all *BmorCSP* protein genes, except *BmorCSP19* and *BmorCSP20*. The degree of mutations was gene-specific. The BmorCSP4 protein isoform was closely related to the native form of the protein (only two replacement sites, the N- and C-termini). BmorCSP7 had only Phe-Val-Phe motif insertion near Cys36. BmorCSP17 had only Gln/Thr, Leu/Phe and Gly/Asn replacements in α-helix α2. All other BmorCSP peptides showed higher level of mutation. In addition, BmorCSP4 BmorCSP7 and BmorCSP17 showed no Gly insertion, but one or two +Gly were found near Cys in all other BmorCSPs.[19,20] Examination of BmorCSP-RNA clones shows numerous sites where AA dinucleotides are replaced by dinucleotides GG, possibly resulting in the substitution of Glutamic Acid or Lysine for Glycine through RNA editing. However, this is not what is happening at the protein level in BmorCSPs, where we found a new Glycine inserted in a native motif of the protein.[19,20] BmorCSP1 and BmorCSP3 showed +Gly at Cys29 and Cys55. BmorCSP2 showed +Gly at Cys36 and Cys55, while BmorCSP11 retained Gly insertion at Cys29 and Cys36. BmorCSP13 and BmorCSP15 had +Gly only at Cys29. BmorCSP6, BmorCSP8, BmorCSP12, BmorCSP14 and BmorCSP15 had +Gly only at Cys55, suggesting that the degree of +Gly mutations and the location of the +Gly both depend on the BmorCSP protein.[19,20] There were no particular mutations in the vicinity of Cysteine residues in BmorCSP1, BmorCSP2, BmorCSP4 and BmorCSP14-RNA clones. In BmorCSP2-encoding RNA clones, many A-to-G substitutions were found in the sequence coding for the motif Lys-Lys-Thr, but none of these substitutions are juxtaposed. They enrol the first and/or the third base on triplets coding for Lys and Thr, respectively, leading to GAG, AGG and GCG tri-nucleotides specifying a switch to Glu, Arg and Ala residues, respectively, but no Gly.[19,20] Finally, genomic DNA and RNA of the mutated proteins were examined for these specific mutations and they are mutation free, strongly suggesting that these mutations including Gly insertion occur at the protein level.[19-23,29]

Most interestingly, +Gly led to removal of specific α-helical maillon (mainly α2).[20] Such a multiplicity in protein isoforms, such a wide set of recombination in numerous amino acid motifs of N-terminus, all these rules following Cys or Gly insertion at crucial sites, their accumulation in the N-terminal arm and their effects on the α-helical structural elements, as well as the tissue-specificity of various functional mutations,[19,20,29] all together support the

existence of totipotent regulation in protein post-translational modifications, at least in the proteasome machinery rising the activity of the pheromone producing cells in the female sex gland of moths (Figure 2).

If RNA editing/protein recombination and expression of a remarkable diversity of mutant molecules are frequently seen in microorganisms/insects as well as in humans, there might be "hot spots" of mutation in the RNA editing mechanism or some genomic sequences with special structure (gene sequence) that are prone to mutations[36,37]. In insects, hot spots are found for mutations associated with insecticide resistance as described for instance in the gene encoding the para-type of voltage-gated sodium channel in moths.[38] In the silkworm, A-to-G mutations are found about every dozen or two dozens base pairs in the RNA encoding CSP. This is quite similar to the C-to-U mutation pattern in *CSPs*. Eventually they can be found juxtaposed as found at position 210-220 in *CSP2* (G-to-A) and at the 5' end of *PBP1* (C-to-U).[19] Mutations sites on CSP and PBP RNA are often present after a GG, a CC or a AA double base pair. However, the mutation rate does vary over *B. mori* CSPs and the remarkably high rates at which A-to-G and C-to-U mutations occur in *BmorCSP14* would suggest that any A or C base could be hot spot of mutation in the RNA sequence.[19] BmorCSP genes have similar percent G and C content (about 39–47 per cent). They also have similar percent A and T content (about 53–61 per cent). In contrast, the frequencies of AA and TT di-nucleotides vary considerably among *BmorCSPs*. In *Bombyx*, CSP genes show also considerable variability in the frequency of the tri-nucleotides GAA, AAG, AAT, ACA, TTC, GTC, ATT and TTT that could be favoured for RNA editing.[19] Most tissues have "private" mutations, but some mutations are not only produced by the gland but also expressed either in the antennae, legs and wings or in the head tissues. Higher mutation level in the intronless CSP gene *BmorCSP14* suggests that the RNA editing process occurred at the RNA level, not at the pre-RNA level.[19] All the mutations may not confer an advantage of evolution, but most of them may confer a new function to the protein. They are non-synonymous, numerous and pinpointed in the coding region of the gene, in crucial elements of the functional protein structure.[19,20,29,39-41]

RNA editing mechanisms in insects and mammals

Only a few mechanisms have been shown to underlie RNA editing either in insects or mammals. Mammals including human expresses three types of ADAR or adenosine deaminases acting on RNA (ADAR1, ADAR2 and ADAR3), while insects have only one, ADAR2. This leaves many unanswered questions concerning the existence of a much more complex tissue-specific multi-RNA editing system.[42-46]

Similarly to *Drosophila*, *B. mori* possesses only one type of ADAR, namely BmorADAR2.[47-48] *BmorADAR2*, double-stranded RNA-specific editase (XM_004925094), is located at gene locus LOC100101209 close to *pancreatic lipase-like enzyme protein* gene (LOC101736798; see scaffold NW_004581724). No further ADAR genes can be found analyzing the silkworm genome. However, three editase transcript variants can be found in the silkworm (XM_012689524, XM_012689525 and XM_012689526), showing 42–44 per cent and 58–70 per cent identity to other mammalian and insect ADAR sequences, respectively. Perhaps transcript diversity in expression of the gene coding for ADAR enzyme indicates that multiple polymerization mechanisms and/or BmorADAR2 variant isoforms are responsible of the multiple variety of post-transcriptional modifications as that observed in *CSPs*. A enzyme such as ADAR type2 is known to dimerize and preferentially to favor its own transcript among a large mixed population of RNA transcripts.[49-50]

In addition of A-to-I base mutations, various *Bombyx* tissues are shown to produce a high number of variant CSP transcripts with distinct structural properties through specific C-to-U and/or U-to-C RNA mechanisms as found for the ion channels in the insect brain tissue.[14,19,51,52] The type of enzyme known to mediate C-to-U base mutation refers to Adenosine Deaminases Acting on tRNA or ADAT enzyme.[53] In the silkworm moth *B. mori,* we find *ADAT* gene at locus LOC101738151 (XM_004933249, XP_004933306; see scaffold NW_004582034). *ADAT* is located on chromosome II in the fruitfly *D. melanogaster*, where it could represent a link between RNA and tRNA editing process.[54] Another mechanism known to perform C-to-U conversion refers to Apolipoprotein B editing catalytic subunit 1 **(**Apobec-1) enzyme functional in specific tissues such as the liver and adipose tissue from vertebrates.[55-57] However, using human APOBEC apolipoprotein B mRNA editing enzyme catalytic subunit 3F (APOBEC3F, XM_011529994) as template in a blastn search against the silkworm database (KAIKObase) did not extract any *APOBEC* gene in the insect, suggesting that insects, such as *Bombyx*, lack of APOBEC activity. Alike *ADAR1* gene, this may illustrate the loss of RNA-dependent RNA polymerization (RdRp) enzymes with high RNA substrate specificity during the course of insect evolution.[58]

Lack of *ADAR1* and *apobec* does not exclude the possibility that insects, such as flies and moths, have developed instead more versatile multispecificity editing enzymes as found in some specific mammalian tissues as well as in yeast

and bacterial organisms.[59,60] Some A-to-I RNA editing enzymes have been shown to be able to convert C-to-U on DNA template.[60] Interestingly, such enzymes can mediate A-to-I conversion only on RNA and C-to-U only on DNA, but not the two types of conversion on the same RNA/DNA strand.[60] This shows that using the same single editing enzyme can possibly lead to a great variety of nucleotide substitutions not only during the process of RNA polymerization, but eventually also during the transcription process from DNA-to-RNA (Figure 3).

DNA-dependent RNA polymerases (DdRps, RNAPs or RNA polymerases) do not make any mistakes during transcription. In the case they do, some specific enzymes, namely RNA polymerase IIs, will be capable of base proofreading and self correction. However, it has been documented that, in certain circumstances, certain DdRps start function as RdRps and lose such self-correction capacity due to change of DNA-to-RNA template.[61-63] Therefore, it is not excluded that, in certain circumstances such as the age, hormonal status, reproductive status and initiation of the pheromone activities, a switch of template in DdRp or RdRp enzyme occurs in specific tissues, leading to a high diversity of nucleotide base mutations as that observed in the sex pheromone gland in sexually mature females of the silkworm moth, *B. mori* (Figure 3).

Ubiquitin, ubiquitination, RNA and protein editing in pheromone gland and totipotent cells

It might be surprising to see such amount of editing pluralities in a tissue such as the pheromone gland, which is apparently devoted to one particular task, i.e., production of the pheromone (Bombykol). To accomplish this task, the gland retains multiple functional processes that are all typical of various systems from cell growth, tissue transformation, synaptic transmission and/or absorption-digestion of fuel molecules in eukaryotic cells: lipid metabolism, endocytosis, exocytosis, immune responses, receptor cascades and on-off control. RNA editing is also relevant for the mammalian nervous system where it highly affects transcripts that code for GABA or glutamate receptors.[64,65] Giving the GABA and glutamate examples helps us understand that editing is crucial to provide a single ion channel with different properties, thereby regulating the activation status of a neuron in the brain. Therefore, RNA and/or protein editing appear to be extremely useful not only for the insect pheromone gland but also in the mammalian neural tissue, apparently bearing only one single function, i.e., production of an odor or transmission of a neural signal.

The high degree of mutations observed in the sex gland and the brain can give us an idea of the amount of editing pluralities that can be expected from tissues of the immune or sensory systems that are devoted to produce proteins for over a million sorts of cognate ligands.[66,67] It seems obvious that RNA and protein trafficking is strongly regulated in a tissue-specific manner. The question of how much tolerance or what is the limit in such a dense intracellular RNA and protein trafficking system may be one key point to address. In particular, the degree of mutations observed in the female moth inquires about the sexual differences in RNA and protein regulation in a given tissue. For instance, there might be marked differences between males and females in the degree of olfactory gene variation mediated via RNA editing in the antennae of insects or the nose of vertebrates. The male nostrils or antennae are usually the receivers of complex odorant signals used as pheromones to facilitate mating[68,69]. As a result, there is strong sexual dimorphism in the moth antennae for instance, with female production of pheromonal compounds used to attract mates.[70] Pheromone communication is known throughout the whole animal kingdom, and even bacteria and plants can have gender-oriented chemical communication systems.[71,72] This may suggest the existence of gender-specific RNA editing mechanisms to regulate chemical communication channels in many kinds of organisms.

If the benefits of perception to one sex or the other are different and involve other modalities than olfaction, then sex differences in other sensory systems could also arise from gender-specific RNA editing. Furthermore, it can be reasonably expected that gender-specific RNA editing occurs in various non-sensory systems. In moths, females enrol not only in production of sexual odour scent, calling and mating, but also in egg maturation, recognition of plant-host and oviposition. Since males and females differ so much in their life history often resulting in increased lifetime in females, females can certainly also display RNA editing mechanisms specifically involved in immune response induction to pathogenic microbes, toxins of plant origin (allelochemicals) and/or insecticide xenobiotic compounds.[73] Such a complex female physiology certainly requires a robust flexible genetic system capable of adjustment to any sudden change in external environment conditions.

This conceit of multi-RNA editing for adaptation certainly applies also to cell differentiation and tissue development. Cell potency is the cell's ability to differentiate into another type of cell. The more cell types a cell can differentiate into, the greater its potency. It might be largely due to a greater

flexibility in the proteasome than was found, for instance, in multi-potent progenitor cells (MPCs).[74] Stem cells are maleable in that they have potential to differentiate into many different tissue lineages. MPCs reside in a niche that keeps the cells in a naïve status until chosen to differentiate. How stem cells and MPCs integrate signals from their environment for quiescence, self-renewal or differentiation in a given tissue largely rely on protein ubiquitylation.[75-77] Ubiquitylation (or ubiquitination) is the enzymatic post-translational mechanism by which ubiquitin is attached to a substrate protein. Ubiquitin is a small protein (8.5 kDa) found in humans and most eukaryotic cells to regulate the fate and process of all other proteins of the tissue body. In the nucleus or the cytoplasm of the "young" cell, ubiquitin plays therefore an essential role in the differentiation of stem cells and MPCs during early tissue development. [75-79]

The function of ubiquitination in relation with RNA editing has been described in various organisms including humans. [78-84] Ubiquitination of RNA polymerases II has been found to regulate DNA damage arrest,[80,81] and ADAR1 enzyme activity has been shown to be involved in hematopoiesis, the genesis of all types of blood cells from formation to differentiation.[82] Propyl isomerase Pin1 and E3 ubiquitin ligase Wwp2 are both known to regulate RNA editing sites by binding to ADAR.[83,84] Additionally, RNA editing and proteasome ubiquitination in particular have been found to govern the stability of core clock components, regulating thereby key mechanisms of post-transcriptional regulation in the circadian clockwork at least in mammals.[85,86] Therefore, alterations of RNA editing and/or ubiquitin pathways can seriously affect a variety of transcripts in many various physiological events, including photoperiodic regulation of gene expression in the pheromone biosynthesis pathway in female moths as well as stem cell self-renewal and survival of differentiating MPCs in human. Disrupting RNA editing properties of the sex pheromone gland might help develop new ways to take on the genetic control of pheromone production, odorant chemical communication and thereby reproduction in a specific insect pest species. This might be an original and very efficient alternative method of insect control. At the same time, controlling RNA editing in stem cells might be a potential breakthrough cure for severe myeloid diseases such as leukemia.[87]

In contrast to totipotent cells (TPCs), MPCs can give rise to multiple but limited types of cell, which could be explained by a more limited editing capacity.[88-98] In *Caenorhabditis elegans*, it has been shown that multiple mechanisms including RNA editing may play a role in maintaining totipotency at different stages of development.[97] Similar studies in fishes and mammals are in further agreement with a role of RNA regulation, gene recoding, microRNAs and RNA binding proteins in determining development differences.[98] Moreover, adult cells from mice can be reprogrammed and eventually retreat to embryonic stem cells, thereby recovering not only totipotency but also RNA editing and specific transcriptomic features.[99] Therefore, totipotency in gene editing through RNA and protein modifications may provide the cell with the ability to differentiate into any cell type then grow into any type of tissue in any conditions. The high multi-level editing of *CSPs* as observed in various insect tissues such as the pheromone gland may just be an illustration of this principle (Figures 3 and 4).

Thus, the insect pheromone gland may be an illustration of the genetic capacity of a totipotent cell characterized by multi-editing of all sorts from A-to-G/C-to-U mutations to Gly/Cys insertion and/or specific substitution of amino acid motifs (Figure 4). Occurence of multi-editing and/or all possible ways of recoding genes suggests a similarity between the insect pheromone gland and the human thymus where multiple A-to-I mutation sites have been described in Alu-containing mRNAs, further expanding the diversity of the self-antigen repertoire in T cells.[100,101] We propose that mutations can happen in RNA/protein level in the thymus as found in the sex pheromone gland (Figure 4). The protein mutations observed in the silkworm were not a result of an artefact. They were specific to a particular tissue, the gland.[19,20] Secondly, the RNA was not sequenced for these particular mutations. RNA clones for BmorCSPs were free from these mutations (Gly insertion and/or addition of a sequence-specific amino acid motif).[19] The genomic DNA was mutation free as well.[19] More efforts were done to prove that RNA-editing machinery was not involved in creating these protein-mutations. Analyzing tons of clones, we never found any triplet or combination of triplets insertion mutation in the RNA sequence. We found some frameshift mutations that completely change the composition of the protein sequence motif, but always in the C-terminus, never in the N-terminus (where the protein mutations accumulate).[19] It is possible that one nucleotide change (such as A-to-I and C-to-U that is caused by regular RNA editing) can result in a frameshift mutation that triggers a major change in the whole protein sequence, but this might not be what has happened in the current situation. At the RNA level, we never found such a drastic change in the composition of the whole protein sequence, except in the C-terminus where the whole tail can change following frameshift and/or stop codon mutations.[19,20]

Peptide mutation and motif recombination largely occurs in the N-terminus of CSPs.[19,20,29] None of these mutations was found in sequences encoding other peptide families in two libraries from the gland.[19] The wild-type protein isoforms were certainly found in the same pool of protein species where the mutations carrying other isoforms were found. Immunoblots show the same protein band corresponding to the wild-type CSP in all tissues investigated, but only the gland showed such a high diversity of peptide mutants in a truncated isoform of CSP.[19] Thus, these mutations discovered in *Bombyx* are not random. They are rather tuned for the protein to have new function.[19,20,29] It might be that RNA editing has happened in the RNA encoding CSP to change one or a few amino acid residues at some crucial sites of the protein structure, a first step of modifications resulting in new isoforms, followed by amino acid changes in these proteins as additional mutations affecting the whole protein sequence, particularly the N-tail. It is a new view of genetic regulation of protein expression through gene/RNA editing followed by series of mutations at the level of protein alone.[23]

RNA and protein editing may be essential for adaptation, defence, reproduction and thereby survival in a constantly changing and often hostile natural environment. Adaptation, evolution and tissue development may not be so different between insects and mammals, including humans. It is worth noting that vertebrate and invertebrate tissue systems show strong similarities in their development, cell production and (epi)genetic control. Insect cells are not stable but can change and evolve, similarly to human/mammalian stem cells. A study on the *Drosophila* midgut refers to the existence of stem cells in the fly gut, similar to what was described in human and mouse.[102] This study is relevant to the insect pheromone gland for two reasons: firstly, fatty acids are extensively metabolized by the gut and the gland and secundly, this may correspond to regeneration of cells in the two tissues.[103,104] While pluripotent cells develop mainly through/via RNA editing,[105] a natural path to multiple diverse functional tissues could involve multi-potent cells characterized by only one main type of RNA editing, as found in the insect brain and/or the human digestive tract[106] (Figure 4). In contrast, a plethora of subtle changes in RNAs and proteins might be crucial for various immune or sensory tissues to respond to a broad diversity of antigenic or signalling stimuli molecules that strongly depend on the surrounding environment.[19,20,34,35] The immune and/or olfactory system needs to recognize and process millions of different chemical stimuli, that may require multiple RNA and protein editing mechanisms at the receptor level[34-35] (Figure 4). Similarly, polymorphism in RNA/protein editing might be strongly necessary to influence wing colour evolution, for instance, in butterfly mimicry.[107] This phenomenon of plethoric amount of changes at both RNA and protein levels is certainly also very important for the cell potency to transform into a specific tissue in a specialized compartment of a complex organismal system.[108] There is, however, no such obvious reason for muscle and/or skin tissue to retain such high propensity in RNA and/or protein editing for the regulation of specific gene expression. Muscle is only involved in locomotion or mastication, therefore mutation amounts can be null without drastically affecting the evolutionary trail (Figure 4).

We propose that a subtle amino acid substitution or complete replacement of protein motifs is a complementary robust modification of the protein in addition of acetylation, glycosylation, methylation and phosphorylation processes, crucial for the regulation of cellular totipotency in insects, spores and zygotes, as well as in higher organisms including humans.[109-112] We bring the genetic conceit of protein synthesis in the ribosome very far from one gene producing one single protein. It is already well established that one gene might be encoded for multiple proteins via alternative splicing or RNA editing. Here, we establish that protein mutations described in the silkworm moth *B. mori* is another source for producing more variants, in addition to epigenetic, alternative splicing and other means of RNA regulation for gene recoding. This new view of genetic mutation goes much beyond Darwin's principle of natural selection. RNA and protein mutations can cause even more useful mutations to occur and selection can have a forward influence on these mutations. The natural environment may not trigger specific changes inside cells, thereby selecting only one positive unharmful key point mutation necessary to promote organism/tissue evolution. Genetic mutations precede any changes in the environment, as a trade-off between transcriptome plasticity, protein diversity and genome evolution as described in Cephalopods, such as squids and octopus.[113] Our results in the female silkworm moth *B. mori* shows that the cell builds mutations of all sorts not only at the RNA level but also at the level of the protein alone, eventually to sense and perfectly blend with the environment. But before all this, these phenomena of pluri-editing in two different compartments of the cell may certainly help build a flexible and multi-functional tool, change the biological properties of protein molecules and thereby reassign their tasks. This might be a crucial prerequisite for biogenesis, development and evolution.

The *cSlo* gene in cochlea hair cells has been shown to induce

and possibly produce via or mediated through alternative splicing about five hundred different protein isoforms, each of them crucial to recognize a specific sound frequency.[114] In the fruit fly *D. melanogaster*, alternative splicing would even allow for the production of about 38,000 different protein isoforms from the gene *Dscam*.[115] Because there are even more possible variants induced by RNA/protein editing, and because a specific tissue such as the moth pheromone gland can use many of these mechanisms to accomplish multiple tasks from lipid metabolism to excretion of a particular odour, we believe that there are not thousands, but millions of different possible combinations leading to new functional protein from a single gene mark. Each totipotent cell may use this complete extensive panoply of all possible combinations from RNA splicing, RNA editing and/or protein re-arrangement to differentiate into a given tissue in a specific physiological system. This may be the conceptual model of a universal phenomenon of adaptation in a changing environment from unicellular protozoa to man through insects and plants.

Conclusion

Insects make crucial models for stem cell research in human. There are some unusual mechanisms for playing with nucleotides and/or amino acids to recompose a specific gene product in the female moth pheromone gland tissue. How the pheromone gland or stem cells will regulate such a dense traffic of RNA transcripts and peptide variants remains to be found. The discovery is profound to understand a basic phenomenon of the eukaryotic cell and rethink therapeutic strategies in genetic diseases.

References

1. Crick F. Central dogma of molecular biology. Nature. 1970;227:561-563.
2. Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA. 1977;74:3171-3175.
3. Chow LT, Gelinas RE, Broker TR, et al. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell. 1977;12:1-8.
4. Gilbert W. The exon theory of genes. Cold Spring Harb Symp Quant Biol. 1987;52:901-905.
5. Gott JM, Emeson RB. Functions and mechanisms of RNA editing. Annu Rev Genet. 2000;34:499-531.
6. Bass BL. RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem. 2002;71:817-846.
7. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem. 2010;79:321-349.
8. Savva YA, Rieder LE, Reenan RA. The ADAR protein family. Genome Biol. 2012;13:252.
9. Xie K, Sowden MP, Dance GS, et al. The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. Proc Natl Acad Sci USA. 2004;101:8114-8119.
10. Conticello SG, Thomas CJ, Petersen-Mahrt SK, et al. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol. 2005;22:367-377.
11. Jacobs MM, Fogg RL, Emeson RB, et al. ADAR1 and ADAR2 expression and editing activity during forebrain development. Dev Neurosci. 2009;31:223-237.
12. Chen Z, Eggerman TL, Bocharov AV, et al. Hypermutation induced by APOBEC-1 overexpression can be eliminated. RNA. 2010;16:1040-1052.
13. Picardi E, Horner DS, Pesole G. Single cell transcriptomics reveals specific RNA editing signatures in the human brain. RNA. 2017;23: 860-865.
14. Song W, Liu Z, Tan J, et al. RNA editing generates tissue-specific sodium channels with distinct gating properties. J Biol Chem. 2004;279:32554-32561.
15. Stapleton M, Carlson JW, Celniker SE. RNA editing in *Drosophila melanogaster*: New targets and functional consequences. RNA. 2006;12:1922-1932.
16. Fukui T, Itoh M. RNA editing in P transposable element read-through transcripts in *Drosophila melanogaster*. Genetica. 2010;138:1119-1126.
17. St Laurent G, Tackett MR, Nechkin S, et al. Genome-wide analysis of A-to-I RNA editing by single molecule sequencing in *Drosophila*. Nat Struct Mol Biol. 2013;20:1333-1339.
18. Li X, Overton IM, Baines RA, et al. The ADAR RNA editing enzyme controls neuronal excitability in *Drosophila melanogaster*. Nucleic Acids Res. 2014;42:1139-1151.
19. Xuan N, Bu X, Liu YY, et al. Molecular evidence of RNA editing in *Bombyx* chemosensory protein family. PLoS ONE. 2014;9:e86932.
20. Xuan N, Rajashekar B, Kasvandik S, et al. Structural components of chemosensory protein mutations in the silkworm moth, *Bombyx mori*. Agri Gene. 2016;2:53-58.
21. Picimbon JF. RNA mutations in the moth pheromone gland. RNA Dis. 2014;1:e240.
22. Picimbon JF. RNA mutations: source of life. Gene Technol. 2014;3:112-122.
23. Picimbon JF. Mutations in the insect transcriptome. J Clin Exp Pathol. 2016;6:3.
24. Liu GX, Ma HM, Xie YN, et al. Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisia tabaci* chemosensory

proteins: role of CSP in insect defense. PLoS ONE. 2016;11:e0154706.
25. Picimbon JF. Biochemistry and evolution of CSP and OBP proteins. *In*: Insect Pheromone Biochemistry and Molecular Biology-The Biosynthesis and Detection of Pheromones and Plant Volatiles. Ed.: G.J. Blomquist, R.G. Vogt, Elsevier Academic Press, London, San Diego, 2003;pp. 539-566.
26. Xuan N, Guo X, Xie HY, et al. Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins. Insect Sci. 2015;22:203-219.
27. Liu GX, Arnaud P, Offmann B, et al. Genotyping and bio-sensing chemosensory proteins in insects. Sensors. 2017;17:1801.
28. Liu GX, Picimbon JF. Bacterial origin of insect chemosensory odor-binding proteins. Gene Transl Bioinform. 2017;3:e1548.
29. Picimbon JF. Evolution of protein physical structures in insect chemosensory systems. *In*: Olfactory concepts of insect control. Ed.: JF. Picimbon, Springer-SBM, Dordrecht, 2017, in press.
30. Picimbon JF, Leal WS. Olfactory soluble proteins of cockroaches. Insect Biochem Mol Biol. 1999;30:973-978.
31. Picimbon JF, Dietrich K, Breer H, et al. Chemosensory proteins of *Locusta migratoria* (Orthoptera : Acrididae). Insect Biochem Mol Biol. 2000a;30:233-241.
32. Picimbon JF, Dietrich K, Angeli S, et al. Purification and molecular cloning of chemosensory proteins from *Bombyx mori*. Arch Insect Biochem Physiol. 2000b;44: 120-129.
33. Picimbon JF, Dietrich K, Krieger J, et al. Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). Insect Biochem Mol Biol. 2001;31:1173-1181.
34. Picimbon JF, Gadenne C. Evolution of noctuid Pheromone Binding Proteins: identification of PBP in the black cutworm moth, *Agrotis ipsilon*. Insect Biochem Mol Biol. 2002;32: 839-846.
35. Abraham D, Löfstedt C, Picimbon JF. Molecular evolution and characterization of pheromone binding protein genes in *Agrotis* moths. Insect Biochem Mol Biol. 2005;35: 1100-1111.
36. Chattopadhyay S, Weissman SJ, Minin VN, et al. High frequency of hotspot mutations in core genes of *Escherichia coli* due to short-term positive selection. Proc Natl Acad Sci USA. 2009;106:12412-12417.
37. Chang MT, Asthana S, Gao SP, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol. 2016; 34: 155-163.
38. Haddi K, Berger M, Bielza P, et al. Identification of mutations associated with pyrethroid resistance in the voltage-gated sodium channel of the tomato leaf miner (*Tuta absoluta*). Insect Biochem Mol Biol. 2012;42:506-513.
39. Lartigue A, Campanacci V, Roussel A, et al. X-ray structure and ligand binding study of a moth chemosensory protein. J Biol Chem. 2002;277:32094-32098.
40. Jansen S, Zídek L, Löfstedt C, et al. 1H, 13C, and 15N resonance assignment of *Bombyx mori* chemosensory protein 1 (BmorCSP1). J Biomol NMR. 2006;36:47.
41. Jansen S, Chmelik J, Zídek L, et al. Structure of *Bombyx mori* Chemosensory Protein 1 in solution. Arch Insect Biochem Physiol. 2007;66:135-145.
42. Maas S, Rich A, Nishikura K. A-to-I RNA editing: recent news and residuals mysteries. J Biol Chem. 2002;278:1391-1394.
43. Schaub M, Keller W. RNA editing by adenosine deaminases generates RNA and protein diversity. Biochimie. 2002;84:791-803.
44. Jantsch MJ, Ohman M. RNA editing by Adenosine Deaminases that Act on RNA (ADARs). *In*: RNA editing. Ed.: U. Goringer, Springer-Verlag, Berlin, Heidelberg, 2008;51-84.
45. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. Nat. Rev. Mol. Cell Biol. 2016, 17, 83-96.
46. Huntley MA, Lou M, Goldstein L, et al. Complex regulation of ADAR-mediated RNA-editing across tissues. BMC Genomics. 2016;17:61.
47. Keegan LP, McGurk L, Palavicini JP, et al. Functional conservation in human and *Drosophila* of Metazoan ADAR2 involved in RNA editing: loss of ADAR1 in insects. Nucl Acids Res. 2011;39:7249-7262.
48. Grice LF, Degnan BM. The origin of the ADAR gene family and animal RNA editing. BMC Evol Biol. 2015;15:4.
49. Dawson TR, Sansam CL, Emeson RB. Structure and sequence determinants required for the RNA editing of ADAR2 substrates. J Biol Chem. 2004;279:4941-4951.
50. Poulsen H, Khandelia H, Morth JP, et al. Neurological disease mutations compromise a C-terminal ion pathway in the Na^+/K^+-ATPase. Nature. 2010;467:99-102.
51. Liu Z, Song W, Dong K. Persistent tetrodotoxin-sensitive sodium current resulting from U-to-C RNA editing of an insect sodium channel. Proc Natl Acad Sci USA. 2004;101:11862-11867.
52. Maas S. RNA editing: current research and future trends. Horizon Scientific Press, 2013;239.

53. Maas S, Gerber AP, Rich A. Identification and characterization of a human tRNA-specific adenosine deaminase related to the ADAR family of pre-mRNA editing enzymes. Proc Natl Acad Sci USA. 1999;96:8895-8900.
54. Keegan LP, Gerber AP, Brindle J, et al. The properties of a tRNA-specific adenosine deaminase from *Drosophila melanogaster* support an evolutionary link between pre-mRNA editing and tRNA modification. Mol Cell Biol. 2000;20:825-833.
55. Chester A, Scott J, Anant S. RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. Biochem Biophys Acta. 2000;1494: 1-13.
56. Gu T, Buaas FW, Simons AK, et al. Canonical A-to-I and C-to-U RNA editing is enriched at 3' UTRs and microRNA target sites in multiple mouse tissues. PLoS ONE 2012;7:e33720.
57. Blanc V, Park E, Schaefer S, et al. Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. Genome Biol. 2014;15:R79.
58. Keegan LP, Gallo A, O'Connell MA. The many roles of an RNA editor. Nat Rev Genet. 2001;2:869-878.
59. Anant S, Davidson NO. Evolutionary origins of the mammalian apolipoproteinB RNA editing enzyme, apobec-1: structural homology inferred from analysis of a cloned chicken small intestinal cytidine deaminase. Biol Chem. 1998; 379:1075-1081.
60. Rubio MAT, Pastar I, Gaston KW, et al. An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. Proc Natl Acad Sci USA. 2007;104:7821-7826.
61. Moraleda G, Taylor J. Host RNA polymerase requirements for transcription of the human hepatitis delta virus genome. J Virol. 2001;75: 10161-10169.
62. McNaughton TB, Shi ST, Modahl LE, et al. Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA polymerases. J Virol. 2002;76:3920-3927.
63. Ahlquist P. RNA-dependent RNA polymerases, viruses, and RNA silencing. Science. 2002;296:1270-1273.
64. Buckingham SD, Biggin PC, Sattelle BM, et al. Insect GABA receptors: splicing, editing, and targeting by antiparasitics and insecticides. Mol Pharmacol. 2005;68:942-951.
65. Rosenthal JJC, Seeburg PH. A-to-I RNA editing: effects on proteins key to neural excitability. Neuron. 2012;74:432-439.
66. Wooldridge L, Ekeruche-Makinde J, van den Berg HA, et al. A single autoimmune T cell receptor recognizes more than a million different peptides. J Biol Chem. 2012;287:1168-1177.
67. Bushdid C, Magnasco MO, Vosshall LB, et al. Humans can discriminate more than 1 trillion olfactory stimuli. Science. 2014;343:1370-1372.
68. Brennan PA. Pheromones and mammalian behavior. In: The Neurobiology of Olfaction, Ed.: A. Menini, Boca Raton (FL): CRC Press/Taylor & Francis, 2010.
69. McNeil JN, Millar JG. Chemical communication: pheromones and allelochemicals. In: Insects: Structure and Function. Author: R.F. Chapman, Ed.: S.J. Simpson, A.E. Douglas, Cambridge University Press, 2012.
70. Sanes JR, Hildebrand JG. Structure and development of antennae in a moth, *Manduca sexta*. Dev Biol. 1976;51:282-299.
71. Bouwmeester HJ, Roux C, Lopez-Raez JA, et al. Rhizosphere communication of plants, parasitic plants and AM fungi. Trends Plant Sci. 2007;12:224-230.
72. Farag M, Zhang H, Ryu CM. Dynamic chemical communication between plants and bacteria through airborne signals: induced resistance by bacterial volatiles. J Chem Ecol. 2013;39:1007-1018.
73. Barthel A, Staudacher H, Schmaltz A, et al. Sex-specific consequences of an induced immune response on reproduction in a moth. BMC Evol Biol. 2015; 15: 282.
74. Ivanovska IL, Shin JW, Swift J, et al. Stem cell mechanobiology: diverse lessons from bone marrow. Trends Cell Biol. 2015;25:523-532.
75. Strikoudis A, Guillamot M, Aifantis I. Regulation of stem cell function by protein ubiquitylation. EMBO Rep. 2014;15:365-382.
76. Naujokat C, Saric T. Concise review: role and function of the ubiquitin-proteasome system in mammalian stem and progenitor cells. Stem Cells. 2007;25:2408-2418.
77. Werner A, Manford AG, Rape M. Ubiquitin-dependent regulation of stem cell biology. Trends Cell Biol. 2017;27:568-579.
78. Nishikura K. Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat Rev Mol Cell Biol. 2006;7:919-931.
79. Shilatifard A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem. 2006;75:243-269.
80. Li H, Zhang Z, Wang B, et al. Wwp2-mediated ubiquitination of the RNA Polymerase II large subunit in mouse embryonic pluripotent stem cells. Mol Cell Biol. 2007;27:5296-5305.
81. Anindya R1, Aygün O, Svejstrup JQ. Damage-induced ubiquitylation of human RNA polymerase II by the ubiquitin ligase Nedd4, but not Cockayne syndrome proteins or BRCA1. Mol Cell. 2007;28:386-397.

82. Hartner JC, Walkley CR, Lu J, et al. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. Nat Immunol. 2009;10:109-115.
83. Li L, Qian G, Zuo Y, et al. Ubiquitin-dependent turnover of ADAR1 is required for efficient antiviral activity of type I-interferon. J Biol Chem. 2016; 291: 24974-24985.
84. Marcucci R, Brindle J, Paro S, et al. Pin1 and WWP2 regulate GluR2 Q/R site RNA editing by ADAR2 with opposing effects. EMBO J. 2011;30:4211-4222.
85. Hirano A, Yumimoto K, Tsunematsu R, et al. FBXL21 regulates oscillation of the circadian clock through ubiquitination and stabilization of cryptochromes. Cell. 2013;152:1106-1118.
86. Terajima H, Yoshitane H, Ozaki H, et al. ADARB1 catalyzes circadian A-to-I editing and regulates RNA rhythm. Nat Genet. 2017;49:146-151.
87. Rossetti C, Picardi E, Ye M, et al. RNA editing signature during myeloid leukemia cell differentiation. Leukemia 2017;doi: 10.1038/leu.2017.134.
88. Hartner JC, Schmittwolf C, Kispert A, et al. Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. J Biol Chem. 2004;279:4894-4902.
89. Iisaza H, Nishikura K. A new function for the RNA-editing enzyme ADAR1. Nat Immunol. 2009;10:16-18.
90. Wright JE, Ciosk R. RNA-based regulation of pluripotency. Trends Genet. 2013;29:99-107.
91. Ye J, Blelloch R. Regulation of pluripotency by RNA binding proteins. Cell Stem Cell. 2014;15:271-280.
92. Emani MR, Närvä E, Stubb A, et al. The L1TD1 protein interactome reveals the importance of post-transcriptional regulation in human pluripotency. Stem Cell Rep. 2015;4:519-528.
93. Hockemeyer D, Jaenisch R. Induced pluripotent stem cells meet genome editing. Cell Stem Cell. 2016;18:573-586.
94. Zipeto MA, Court AC, Sadarangani A, et al. ADAR1 activation drives leukemia stem cell self-renewal by impairing Let-7 biogenesis. Cell Stem Cell. 2016;19:177-191.
95. Baysal BE, Sharma S, Hashemikhabir S, et al. RNA editing in pathogenesis of cancer. Cancer Res. 2017;77:1-7.
96. Jiang Q, Crews LA, Holm F, et al. RNA editing-dependent epitranscriptome diversity in cancer stem cells. Nat Rev Cancer. 2017;17:381-392.
97. Ciosk R, Depalma M, Priess JR. Translational regulators maintain totipotency in the *Caenorhabditis elegans* germline. Science. 2006;311:851-853.
98. Kedde M, Agami R. Interplay between microRNAs and RNA-binding proteins determines developmental processes. Cell Cycle. 2008;7:899-903.
99. Abad M, Mosteiro L, Pantujo C, et al. Reprogramming in vivo produces teratomas and iPS cells with totipotency features. Nature. 2013;502:340-345.
100. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2004;2:e391.
101. Danan-Gotthold M, Guyon C, Giraud M, et al. Extensive RNA editing and splicing increase immune self-representation diversity in medullary thymic epithelial cells. Genome Biol. 2016;17:219.
102. Ohlstein B, Spradling A. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. Nature. 2006;439:470-474.
103. Ito K, Suda T. Metabolic requirements for the maintenance of self-renewing stem cells. Nat Rev Mol Cell Biol. 2014;15:243-256.
104. Chandel NS, Jasper H, Ho TT, et al. Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. Nat Cell Biol. 2016;18:823-832.
105. Choi YJ, lin CP, Risso D, et al. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. Science. 2017;355:eaag1927.
106. Maas S, Rich A. Changing genetic information through RNA editing. BioEssays. 2000;22:790-802.
107. Thompson MJ, Timmermans MJ, Jiggins CD, et al. The evolutionary genetics of highly divergent alleles of the mimicry locus in *Papilio dardanus*. BMC Evol Biol. 2014;14:140.
108. Baumann K. Stem cells A key to totipotency. Nat Rev Mol Cell Biol. 2017;18:137.
109. Morgan TH. Regeneration. London: MacMillan & Co. Ltd., 1901, p. 242.
110. Corley LS, Lavine MD. A review of insect stem cell types. Semin Cell Dev Biol. 2006;17:510-517.
111. Potrykus I. From the concept of totipotency to biofortified cereals. Annu Rev Plant Biol. 2015;66:1-22.
112. Wang YC, Peterson SE, Loring JF. Protein post-translational modifications and regulation of pluripotency in human stem cells. Cell Res. 2014;24:143-160.
113. Liscovitch-Brauer N, Alon S, Porath HT, et al. Trade-off between transcriptome plasticity and genome evolution in cephalopods. Cell. 2017; 169:191-202.
114. Black DL. Splicing in the inner ear: a familiar tune, but what are the instruments? Neuron. 1998;20:165-168.
115. Neves G, Zucker J, Daly M, et al. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. Nature Genet. 2004;36:240-246.

Not commissioned. Externally peer reviewed.

ACKNOWLEDGEMENTS

Overseas high-level talent program, Taishan scholarship and Shandong Province.

FUNDING

NO.tshw20091015, ZR2011CM046

PEER REVIEW

Figure 1: Sex pheromone communication in the silkworm moth, *Bombyx mori*. (A) Sexually-mature *B. mori* female in calling posture: stretched wings, arisen abdomen and pheromone-producing gland (PG) largely exposed at the tip of the abdomen. (B) Male is attracted from a distance by a specific pheromonal compound (bombykol) that mediates odour interaction between the female and the male of adult moths, prelude to encountering, mating behaviour and thereby reproduction.

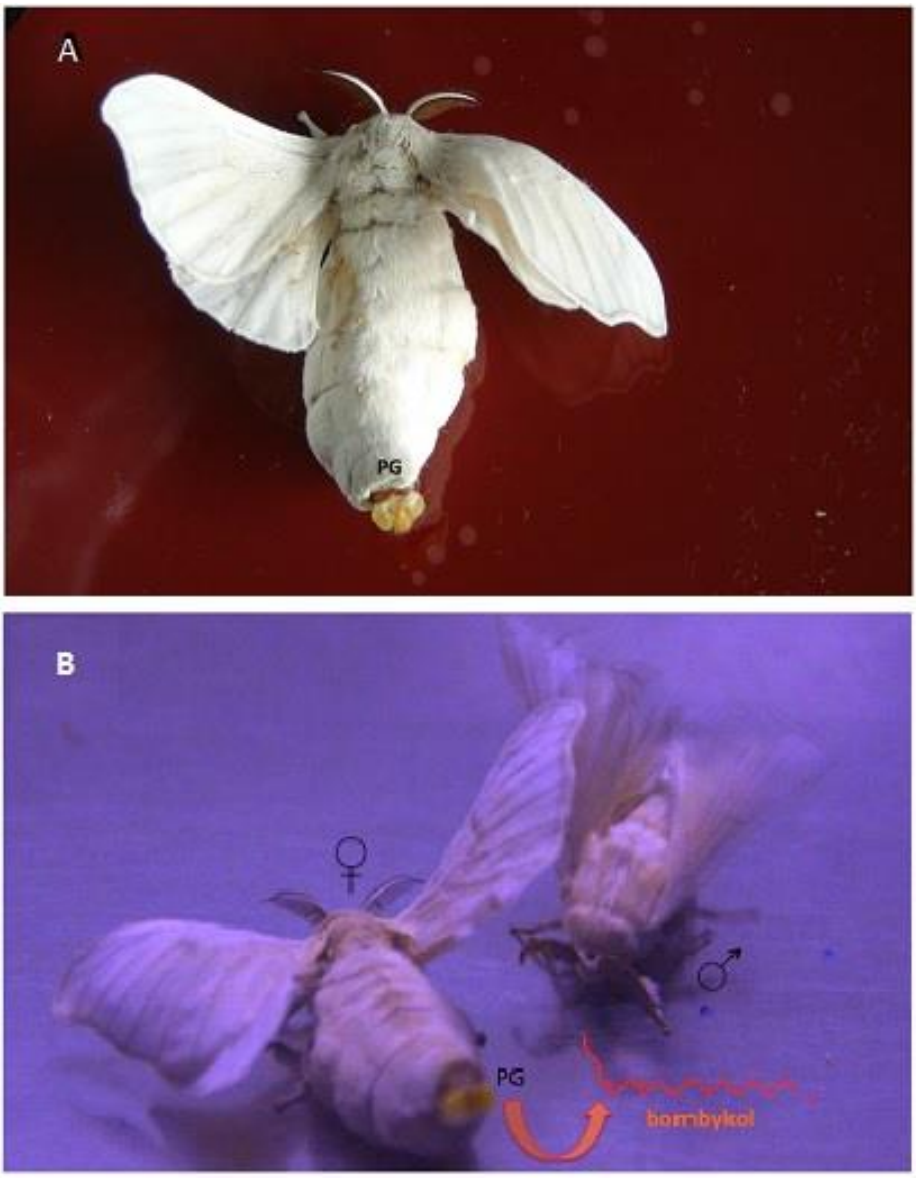

Figure 2: Specific mutational series at both RNA and protein levels in CSPs from the moth pheromone gland. A) Microscopic view of the silkworm moth pheromone gland (PG) sandwiched between sclerotized ovipositor valves (sov) and sclerotized cuticle that is mostly removed. B) Ratio of nucleotide mutations identified at the RNA level in various moth tissues. ADAR: adenosine deaminase acting on RNA; Apobec: apolipoprotein B editing catalytic subunit; RNA pol II: RNA polymerase II; DdRp: DNA-dependent RNA polymerase. C) Glycine insertion mutations identified at the protein level in mixtures of CSPs from the moth pheromone gland. Red triangle indicates the position of Glycine insertion near a specific Cysteine residue in the primary amino acid sequence of the CSP protein. The number above Cys indicates the position of the Cysteine residue in the protein structure. In green: amino acid substitution mutations in the CSP peptide; 2 residues (Ala-Lys) changed to 1 (Gln). In blue: insertion of a Phe-Val-Phe motif instead of Glycine residue at the flank of Cysteine at position 36 (BmorCSP7). [19-20]

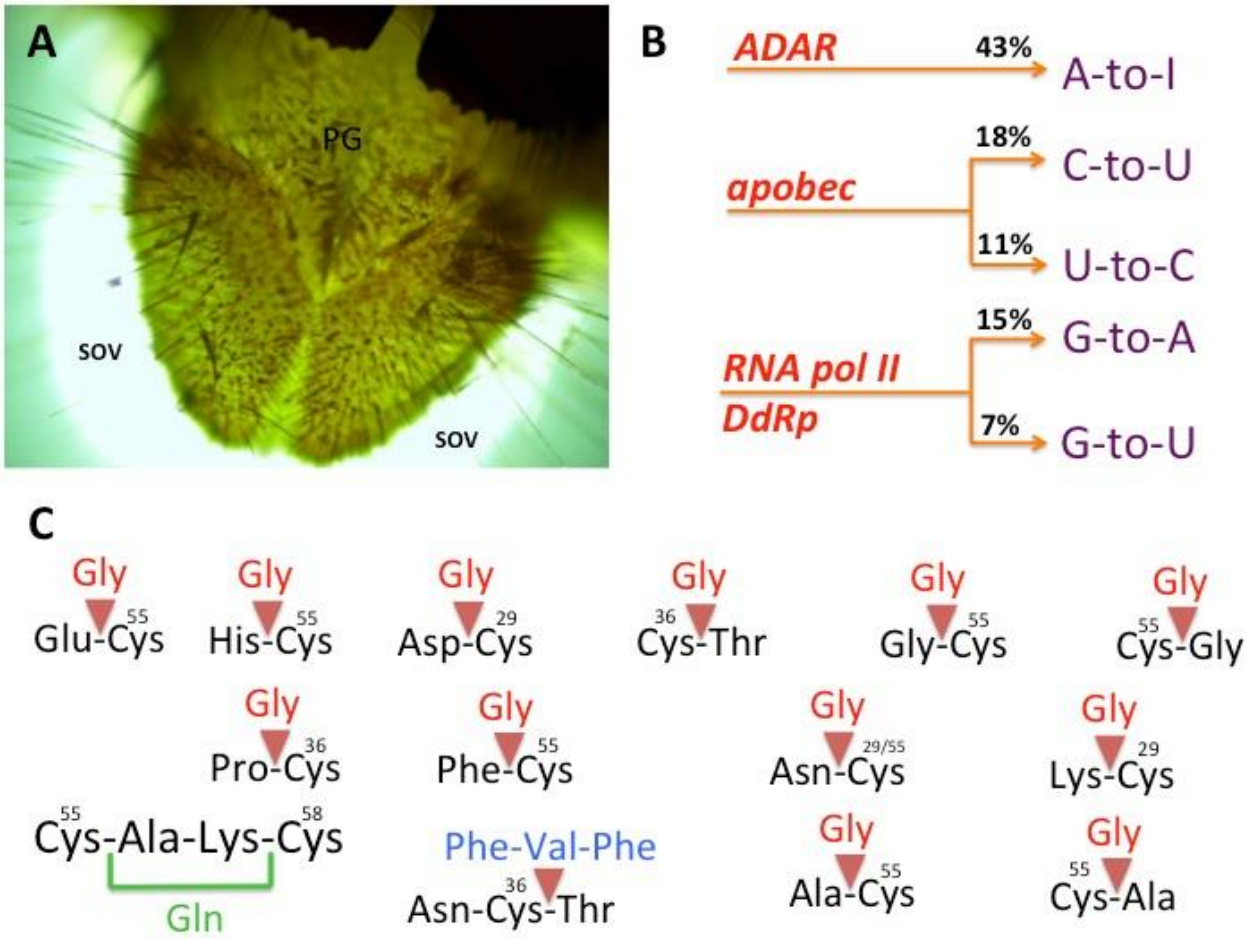

Figure 3: RNA and protein editing-driven expression of *CSPs*. Initial production of new protein motifs can occur through DNA/RNA-dependent RNA polymerization. A CSP protein with the motif Met-Asp-Cys encoded by the three triplets ATG-GAT-TGC leads to multiple variants through RNA polymerase and/or RNA editing systems. A DNA-dependent RNA polymerase (DdRp) introduces site-specific T-to-C, G-to-A and G-to-U mutations into the RNA nucleotide sequence. During transcription, RNA polymerase (RNA pol II) synthesizes the formation of a faithful copy of the DNA template together with a few RNA sequence variants, differing by subtle pinpointed base-pair mutations (1). Native and mutant sequences are then subjects to a high level of RNA editing mediated by RNA-dependent RNA polymerases (RdRps; ADAR, ADAT and/or apobec-*like*), leading to an increased number of RNA variants eventually bearing two or many more nucleotide replacements (2). Mutant RNA strands are therefore translated into amino acid chains corresponding to protein variants either with only one or two amino acid substitutions, enough to cause a large-scale conformational switch, or with a number of completely different amino acid motifs. (3) The native protein sequence and the protein variant isoforms are the templates for further editing in the proteasome. Protein mutations involve not only Glycine, Alanine or multiple residues insertion (+), but also the replacement of specific residue motifs by a single amino acid residue (Gln), a pair of amino acids (Gln-Glu, Glu-Pro or Pro-Glu) or a new completely different amino acid motif. In the added new motif, there can be multiple combinations of two, three or four different amino acids, in order to increase the variance of the protein at such a crucial location. The combination schemes can be different depending on the motif. In the added motif Ala-Cys-Thr-Lys (BmorCSP6 isoform), Ala-Cys converts to Cys-Ala; the order and the position of Thr-Lys do not change. The mixed arrows indicate additional exchange of single amino acids for new functional groups in the protein variant. Before or after conversion mutation, Glycine insertion may occur increasing further the number of mutant variants at this specific location. Finally, a complete amino acid motif can be replaced by one single Cysteine residue, possibly resulting in the production of new disulfide bridges in the functional structure of the protein (BmorCSP11) through specific protein editing pathways.[19,20] DdRp: DNA-dependent RNA polymerase (DdRp), RNApol: RNA polymerase, RdRp: RNA-dependent RNA polymerase (RdRp), ADAR: Adenosine Deaminase Acting on RNA (A-to-I), APOBEC-*like*: Apolipoprotein B mRNA-editing enzyme-catalytic-*like* (C-to-U).

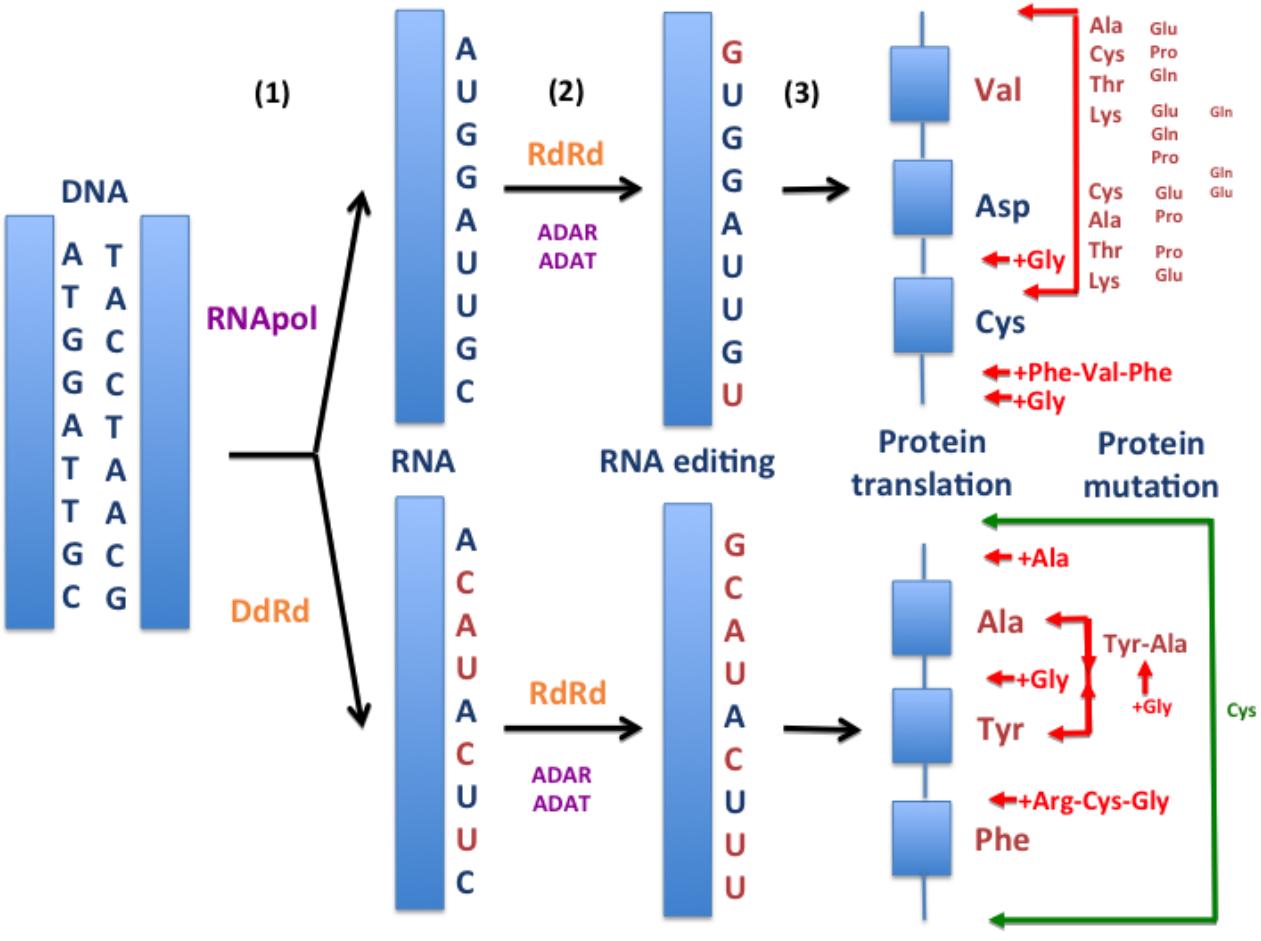

Figure 4: Evolution of totipotent cells to adult tissues through RNA and protein editing. A totipotent cell (TPC) is a particularly unstable cell that is capable of developing into any daughter offspring secondary cell type. A totipotent stem cell can therefore produce every cell type from the embryo and extra-embryonic tissues. During embryo development or cell tissue regeneration, a totipotent cell transforms into a pluripotent stem cell (PPC) (1) that develops into a multipotent cell (MPC) (2) and further on into a particular tissue of a complete organ in a complex organism, such as insect, mammal or human (3-3'). Specific development of TPCs, PPCs and MPCs would be due to the cell's differential capacity for both RNA and protein editing. TPCs would be characterized by high levels of editing in both RNA and protein, while PPCs would develop only thanks to RNA editing. The level of RNA editing would decrease with further tissue genesis. MPCs would conserve either A-to-I or C-to-U/U-to-C mutation capacity, depending on the type of tissue in which they aim to form a specific function. A-to-I, C-to-U and U-to-C are nucleotide base conversions from some known RNA editing enzymes (ADAR, ADIT and apobec). "…" in red indicates additional series of base conversion (U-to-A, U-to-G, G-to-A, G-toU and/or C-to-I) that may occur at the level of the DNA transcription into RNA strands (DNA-dependent RNA polymerization). The bold arrow shows a direct transition from TPC and PPC to pheromone gland and wing tissues, respectively, while the gut and fat body (liver) as well as the brain and the nervous system originate from MPC. Sensory epithelium and specialized primary lymphoid organs of the immune system (i.e., thymus) would keep a RNA and protein editing system functionally similar to TPCs. Specifically, mutation-derived peptides would only occur in TPCs and some types of tissues such as the insect pheromone gland and the human/mammalian thymus. Similarly to TPCS and immune T cells, sensory receptor cells in the nose or the insect antennae would retain high RNA and protein editing capacity to perceive and respond to a million of different stimuli (e.g., ligands), while some more neutral tissues such as muscle and epidermis or skin would loose it. +Cys: Cysteine insertion, +Gly: Glycine insertion, Gln: replacement of specific amino acid motif by Glutamine residue. "…" in blue indicates multiple series of amino acid residue mutations at the protein level alone as discovered in the pheromone gland of the silkworm moth, *B. mori*. [19-20]