Small changes to DNA that were once considered innocuous enough to be ignored are proving to be important in human diseases, evolution and biotechnology

By J. V. Chamary and Laurence D. Hurst

Biologists long thought they understood how genetic mutations cause disease. But recent work has revealed an important twist in the tale and uncovered surprising—even counterintuitive—ways that alterations in DNA can make people sick. The classic view assumed that what are termed “silent” mutations were inconsequential to health, because such changes in DNA would not alter the composition of the proteins encoded by genes. Proteins function in virtually every process carried out by cells, from catalyzing biochemical reactions to recognizing foreign invaders. Hence, the thinking went, if a protein’s makeup ends up being correct, any small glitches in the process leading to its construction could not do a body harm.

Yet detective work occasionally traced a disorder to a silent mutation, even though researchers presumed that it could not possibly be the culprit. Similar mysteries popped up in studies of genome evolution, where patterns of changes in the DNA of various species indicated that many silent mutations were preserved over time—a sign that they were useful to the organisms possessing them. In many species, these changes seemed to help cells make proteins more efficiently, but not in people.

Now investigators are beginning to tease out the effects that silent mutations can have on human health and disease. And findings are suggesting intriguing new avenues for improving the design of genes meant to be used as therapies and for genetic engineering.

**Synonymous but Not the Same**

How a gene mutation can leave a protein unaffected becomes clear when one looks at the way cells make proteins. The basic formula is simple: a string of DNA nucleotides gives rise to a nearly identical sequence of RNA nucleotides, which in turn is translated into an amino acid chain that folds itself neatly into a protein. The letters of this nucleic acid alphabet are distinguished from one another by their chemical bases—adenine (A), cytosine (C), guanine (G) and thymine (T) in DNA, with uracil (U) substituting for thymine in RNA. In other words, the instructions encoded in nucleic acids must be converted into the language of amino acids so that their “meaning” (a useful protein) can be expressed. When a gene is thus “expressed,” the strands of the DNA double helix separate and cellular machinery transcribes the nucleotide sequence along a single strand into a copy made of RNA. Then this messenger RNA (mRNA) transcript must often be edited into a briefer form before it is ready to be translated into a protein by ribosomes and smaller RNAs called transfer RNA (tRNA). As ribosomes ride along the mRNA, tRNAs arrive to deliver the encoded amino acids. Each tRNA carries a specific amino acid, and most are able to recognize just one particular three-nucleotide sequence in the mRNA strand. When a tRNA meets its mRNA match, the ribosome adds the tRNA’s amino acid cargo to the growing amino acid chain [see box on page 49].
The code that cells use to translate the language of DNA and RNA into protein—the famous genetic code—is thus merely the set of rules that govern which tRNA bears which amino acid. This code has a critical feature: it is redundant. All genes and their mRNAs are organized into three-letter “words,” called codons. Sixty-four three-letter codons can be constructed from the four-nucleotide alphabet. Three codons act as stop signals to halt RNA translation, which still leaves 61 possible codons to specify a protein alphabet of only 20 amino acids, so nearly every amino acid is specified by more than one codon. For instance, all codons starting with GG (GGA, GGC, GGG, GGU) are translated to the amino acid glycine, making those codons synonyms.

Single-letter changes to the DNA, known as point mutations, can therefore change a codon to one that specifies the wrong amino acid (known as a missense mutation) or to a stop signal (nonsense mutation), causing the final protein to be truncated. A single-base change can also alter a stop codon so that it then encodes an amino acid (sense mutation), resulting in a lengthened protein. And a final change is possible: a mutation that alters a nucleotide but yields a synonymous codon. These mutations are the ones termed “silent.”

Evidence of Bias

Examples certainly abound of the first three types of point mutations having a major impact on human health. Three different point mutations in the genes encoding proteins that make up the hemoglobin molecules in red blood cells are responsible for three separate and grave diseases, for instance. In the case of sickle cell anemia, a missense mutation exchanges a water-loving (hydrophilic) amino acid for a water-avoiding one (hydrophobic), causing the proteins to clump together and produce characteristic sickle-shaped blood cells. In polycythemia disorders, a nonsense mutation truncates one of the hemoglobin proteins, resulting in thickened blood. And in thalassemia, a sense mutation changes a stop codon (TAA) to the codon for glutamine (CAA), creating a much longer and nonfunctional protein.

Only in the 1980s did scientists realize that silent mutations could also affect protein production—at least in bacteria and yeast. A key discovery at the time was that the genes of those organisms did not use synonymous codons in equal numbers. When the bacterium Escherichia coli specifies the amino acid asparagine, for instance, the codon AAC appears in its DNA much more often than AAT. The reason for this biased usage of codons soon became apparent: cells were preferentially employing certain codons because those choices enhanced the rate or accuracy of protein synthesis.

It turned out that tRNAs corresponding to those synonymous codons typically are not equally abundant within the cell. Most important, then, a gene that contains more of the codons matching the relatively abundant tRNAs would be translated faster, because the higher concentration of those tRNAs would make them more likely to be present when needed. In other cases, a single tRNA variety matches more than one synonymous codon but binds more readily to one codon in particular, so the use of that codon maximizes the accuracy of translation. Consequently, a cell has good reasons not to use all codons equally. As expected, in bacteria and yeast the genes that encode especially abundant proteins exhibit the greatest codon bias, with the preferred codons matching the most common or better-binding tRNAs.

Later observations in other organisms—including plants, flies and worms—revealed similar biases. With such a diverse array of species employing this technique to improve the efficiency of protein production, it seemed likely that mammals would, too. Analyses of mammalian genes did indeed reveal tendencies toward favoring certain codons. The similarity between simple organisms and mammals, however, proved to be only superficial. For reasons not yet fully understood, mammalian genomes are organized into large blocks, each with a distinctively skewed nucleotide content: some regions are rich in G and C bases, whereas others are enriched for A and T. As a result, genes residing in a GC-rich region of the genome tend to have many codons containing those bases. Our genes, then, do show a bias for using certain codons, but unlike simpler organisms, the mammalian pattern does not obviously suggest that the reason is to optimize protein synthesis.

For many years these findings seemed to diminish the likelihood that silent mutations influenced the functioning of the human body. Starting in the early 2000s, however, comparisons of the same gene in different species began to hint that this orthodoxy was wrong. One can measure the rate at which gene sequences in two species have diverged by comparing the sites where nucleotides have changed and those cases, a missense mutation exchanges a water-loving (hydrophilic) amino acid for a water-avoiding one (hydrophobic), causing the proteins to clump together and produce characteristic sickle-shaped blood cells. In polycythemia disorders, a nonsense mutation truncates one of the hemoglobin proteins, resulting in thickened blood. And in thalassemia, a sense mutation changes a stop codon (TAA) to the codon for glutamine (CAA), creating a much longer and nonfunctional protein.

Evidence of Bias

Examples certainly abound of the first three types of point mutations having a major impact on human health. Three different point mutations in the genes encoding proteins that make up the hemoglobin molecules in red blood cells are responsible for three separate and grave diseases, for instance. In the case of sickle cell anemia, a missense mutation exchanges a water-loving (hydrophilic) amino acid for a water-avoiding one (hydrophobic), causing the proteins to clump together and produce characteristic sickle-shaped blood cells. In polycythemia disorders, a nonsense mutation truncates one of the hemoglobin proteins, resulting in thickened blood. And in thalassemia, a sense mutation changes a stop codon (TAA) to the codon for glutamine (CAA), creating a much longer and nonfunctional protein.

Only in the 1980s did scientists realize that silent mutations could also affect protein production—at least in bacteria and yeast. A key discovery at the time was that the genes of those organisms did not use synonymous codons in equal numbers. When the bacterium Escherichia coli specifies the amino acid asparagine, for instance, the codon AAC appears in its DNA much more often than AAT. The reason for this biased usage of codons soon became apparent: cells were preferentially employing certain codons because those choices enhanced the rate or accuracy of protein synthesis.

It turned out that tRNAs corresponding to those synonymous codons typically are not equally abundant within the cell. Most important, then, a gene that contains more of the codons matching the relatively abundant tRNAs would be translated faster, because the higher concentration of those tRNAs would make them more likely to be present when needed. In other cases, a single tRNA variety matches more than one synonymous codon but binds more readily to one codon in particular, so the use of that codon maximizes the accuracy of translation. Consequently, a cell has good reasons not to use all codons equally. As expected, in bacteria and yeast the genes that encode especially abundant proteins exhibit the greatest codon bias, with the preferred codons matching the most common or better-binding tRNAs.

Later observations in other organisms—including plants, flies and worms—revealed similar biases. With such a diverse array of species employing this technique to improve the efficiency of protein production, it seemed likely that mammals would, too. Analyses of mammalian genes did indeed reveal tendencies toward favoring certain codons. The similarity between simple organisms and mammals, however, proved to be only superficial. For reasons not yet fully understood, mammalian genomes are organized into large blocks, each with a distinctively skewed nucleotide content: some regions are rich in G and C bases, whereas others are enriched for A and T. As a result, genes residing in a GC-rich region of the genome tend to have many codons containing those bases. Our genes, then, do show a bias for using certain codons, but unlike simpler organisms, the mammalian pattern does not obviously suggest that the reason is to optimize protein synthesis.

For many years these findings seemed to diminish the likelihood that silent mutations influenced the functioning of the human body. Starting in the early 2000s, however, comparisons of the same gene in different species began to hint that this orthodoxy was wrong. One can measure the rate at which gene sequences in two species have diverged by comparing the sites where nucleotides have changed and those cases, a missense mutation exchanges a water-loving (hydrophilic) amino acid for a water-avoiding one (hydrophobic), causing the proteins to clump together and produce characteristic sickle-shaped blood cells. In polycythemia disorders, a nonsense mutation truncates one of the hemoglobin proteins, resulting in thickened blood. And in thalassemia, a sense mutation changes a stop codon (TAA) to the codon for glutamine (CAA), creating a much longer and nonfunctional protein.

Evidence of Bias

Examples certainly abound of the first three types of point mutations having a major impact on human health. Three different point mutations in the genes encoding proteins that make up the hemoglobin molecules in red blood cells are responsible for three separate and grave diseases, for instance. In the case of sickle cell anemia, a missense mutation exchanges a water-loving (hydrophilic) amino acid for a water-avoiding one (hydrophobic), causing the proteins to clump together and produce characteristic sickle-shaped blood cells. In polycythemia disorders, a nonsense mutation truncates one of the hemoglobin proteins, resulting in thickened blood. And in thalassemia, a sense mutation changes a stop codon (TAA) to the codon for glutamine (CAA), creating a much longer and nonfunctional protein.

Only in the 1980s did scientists realize that silent mutations could also affect protein production—at least in bacteria and yeast. A key discovery at the time was that the genes of those organisms did not use synonymous codons in equal numbers. When the bacterium Escherichia coli specifies the amino acid asparagine, for instance, the codon AAC appears in its DNA much more often than AAT. The reason for this biased usage of codons soon became apparent: cells were preferentially employing certain codons because those choices enhanced the rate or accuracy of protein synthesis.

It turned out that tRNAs corresponding to those synonymous codons typically are not equally abundant within the cell. Most important, then, a gene that contains more of the codons matching the relatively abundant tRNAs would be translated faster, because the higher concentration of those tRNAs would make them more likely to be present when needed. In other cases, a single tRNA variety matches more than one synonymous codon but binds more readily to one codon in particular, so the use of that codon maximizes the accuracy of translation. Consequently, a cell has good reasons not to use all codons equally. As expected, in bacteria and yeast the genes that encode especially abundant proteins exhibit the greatest codon bias, with the preferred codons matching the most common or better-binding tRNAs.

Later observations in other organisms—including plants, flies and worms—revealed similar biases. With such a diverse array of species employing this technique to improve the efficiency of protein production, it seemed likely that mammals would, too. Analyses of mammalian genes did indeed reveal tendencies toward favoring certain codons. The similarity between simple organisms and mammals, however, proved to be only superficial. For reasons not yet fully understood, mammalian genomes are organized into large blocks, each with a distinctively skewed nucleotide content: some regions are rich in G and C bases, whereas others are enriched for A and T. As a result, genes residing in a GC-rich region of the genome tend to have many codons containing those bases. Our genes, then, do show a bias for using certain codons, but unlike simpler organisms, the mammalian pattern does not obviously suggest that the reason is to optimize protein synthesis.

For many years these findings seemed to diminish the likelihood that silent mutations influenced the functioning of the human body. Starting in the early 2000s, however, comparisons of the same gene in different species began to hint that this orthodoxy was wrong. One can measure the rate at which gene sequences in two species have diverged by comparing the sites where nucleotides have changed and those
The genetic code, which governs how a cell translates DNA instructions, via RNA, into functional proteins, is unusual in that it is redundant. Genes "written" in RNA nucleotides spell out the sequence of amino acids in an encoded protein using three-letter words called codons that correspond to one of 20 amino acids (table). With an alphabet of four nucleotide bases, 64 codon triplets are possible—resulting in several codons that specify the same amino acid. A DNA mutation that changes one of these codons to its synonym should therefore be "silent" in protein terms.

**TRANSCRIPTION AND EDITING**

Inside the cell nucleus, the DNA double helix unwinds to allow an RNA copy of a gene to be made. The resulting transcript is then edited to remove segments that do not encode amino acids, producing a shorter messenger RNA (mRNA) version. Pairing of the bases in the RNA nucleotides causes the mRNA molecule to adopt a folded structure.

**THE CODON–AMINO ACID CODE**

Because the four RNA bases (A, C, G, U) yield 64 possible triplet combinations, more than one codon can specify a particular amino acid. Often such synonymous codons differ only in their third nucleotide positions.

**BREAKING THE SILENCE**

Initially researchers had no idea how such mutations could disturb protein manufacture in mammals. Lately, however, studies of human disease have provided not just one mechanism but many. Silent disease-causing mutations interfere with several stages of the protein-making process, from DNA transcription all the way through to the translation of mRNA into proteins.

One example involves silent mutations changing how a gene transcript is edited. Shortly after a gene is transcribed into RNA form,
Synonymous codons may specify the same amino acid, but a mutation that changes one codon to its synonym can alter a gene's encoded message if it interferes with the cell's editing of mRNA. Many diseases are caused by such editing errors, and a gene involved in cystic fibrosis illustrates how even so-called silent mutations can cause a gene's protein meaning to change (bottom).

**NORMAL RNA SPlicing**
The raw RNA transcript of a gene contains exons, which encode amino acids, and long noncoding intron segments that must be edited out of the final mRNA. Within each exon, short nucleotide sequences act as exonic splicing enhancers (ESE) that flag the boundaries of the exon to cellular editing machinery. The binding of splicing regulatory (SR) proteins to enhancer sites directs "spliceosome" proteins to both ends of an intron, which they excise from the transcript, before joining the exon ends together.

**EXON SKIPPING**
Single-nucleotide synonymous changes in an exon can render splicing enhancer sequences invisible to the splicing machinery, causing an entire exon to be left out of the final mRNA.

**PROTEIN ALTERING**
Mutations in the cystic fibrosis transmembrane-conductance receptor (CFTR) gene that disable the receptor protein are implicated in cystic fibrosis and several other related disorders. In an experiment to test whether silent mutations could also affect the CFTR protein, scientists induced single-nucleotide mutations, one by one, to create synonymous codons in CFTR exon 12, then analyzed the resulting proteins. The six synonymous mutations shown (one quarter of those tested) each caused exon 12 to be skipped during mRNA editing, yielding a truncated CFTR protein.

That transcript is trimmed to remove noncoding regions known as introns. Like a movie editor who cuts out unwanted film, cellular splicing machinery needs to find the good bits that encode amino acids, known as exons, and then splice them together to produce the final mRNA version of the gene. Human genes are especially rich in introns, with each gene having an average of eight long intronic stretches, so the splicing machinery needs a way to tell where each exon starts and ends.

Research over the past few years has revealed that exons not only specify amino acids, they also contain within their sequences cues necessary for intron removal. Chief among these are exonic splicing enhancer (ESE) motifs—short sequences of about three to eight nucleotides that sit near the ends of the exons and define the exon for the cellular splicing machinery. The need for such motifs can in fact explain a preference for certain nucleotides in human genes. Although the codons GGA and GGG, which encode glycine, can both occur in splicing enhancers, the former codon acts as a more potent enhancer, leading to more efficient splicing. GGA is also correspondingly more common close to the ends of exons.

In support of the view that preserving codon sequence in splicing enhancers matters, research we did with our former University of Bath colleague Joanna L. Parmley has shown that exonic motifs that apparently function as splicing enhancers show slower evolution in their synonymous codons than do neighboring sequences uninvolved in splicing. This slow evolution indicates that natural selection has kept enhancer motifs relatively unchanged because their specific sequences are so significant. Silent altera-
tions to codons containing these enhancers, although they do not change an amino acid, can nonetheless have a major effect on a protein simply because they disrupt the proper removal of introns.

Indeed, when William Fairbrother, now at Brown University, and his colleagues in Christopher Burge’s laboratory at the Massachusetts Institute of Technology compared the ends of exons, they found that people are rather similar to one another. These splice-associated regions lack much variation, even at sites where a mutation would be silent. The reason is not that mutations are not happening at the ends of our exons, but that when they appear, the mutations are so damaging to protein production that they tend to be disruptive and disappear from the population.

To date, some 50 genetic disorders have been linked to silent mutations, many of which also appear to interfere with intron removal. Splicing enhancers can overlap with a considerable length of a gene’s protein-coding sequence, imposing significant limitations on where a silent mutation would be tolerated. A striking example of the damage a mutation in a splicing enhancer can do was recently documented by Francisco Baralle of the International Center for Genetic Engineering and Biotechnology in Trieste, Italy. The investigators found that 2.5 percent of the silent mutations they induced in one exon of the cystic fibrosis transmembrane-conductance regulator (CFTR) gene disrupted splicing and presumably would thus contribute to cystic fibrosis or related disorders.

That is not to say that disruption of splicing is the only mechanism by which silent mutations can cause disease. Even if the introns are correctly removed from an RNA transcript, the mRNA might not fold properly. Contrary to what is sometimes (for simplicity) depicted in textbooks, an mRNA is not just an unstructured linear string. Like the nucleotide pairs that form between the two DNA strands, separate regions of an mRNA can be complementary and will pair up to create an intricate folded structure known as a stem-loop. The way an mRNA folds determines its stability, which in turn can affect the speed of its translation by ribosomes as well as its subsequent degradation by cellular garbage disposals.

In the dopamine receptor D2 gene, which encodes a cell-surface receptor that detects the presence of the neurotransmitter dopamine, one silent mutation causes the mRNA to be degraded more rapidly than normal. As a result, less of the encoded protein is made, leading to cognitive disorders. Conversely, in the catechol-O-methyltransferase (COMT) gene, a silent mutation increases the extent of mRNA folding, possibly creating too much structure that may be hard to unpack before translation—lowering protein synthesis. Andrea G. Nackley and her colleagues at the University of North Carolina at Chapel Hill found that this mutation affects pain tolerance; perhaps it is no surprise that this research was done in a dental school.

Another instance of a silent mutation affecting a protein is also among the most direct effects and involves a gene known as multidrug resistance 1. The gene is so named because its protein product is a cellular pump that in cancer cells helps to expel chemotherapy drugs, thus conferring drug resistance on those cells. Chava Kimchi-Sarfaty and her colleagues at the National Cancer Institute found that the silent change caused the pump protein to misfold, reducing cells’ ability to eject drugs. Because the translation process and protein folding can occur simultaneously, the researchers theorized that the rarer synonymous codon produced by the silent mutation caused a pause during translation, which in turn allowed the protein time to adopt an unusual structure. If this pause occurs, the precise cause is unclear and can be added to the list of as yet unsolved mysteries about the workings of genes and proteins.

**APPLICATION**

### Biased Vaccine

Manipulating the sites of synonymous mutations has allowed scientists to design genes that speed up protein manufacture, but the same technique can also be used to slow it down. Steffen Mueller and his colleagues at Stony Brook University recently took this approach to design a safer polio vaccine. Live viruses make the most potent vaccines because they provoke a strong immune response in the recipient, but live vaccines can reproduce and mutate, potentially causing disease. Mueller’s group took advantage of microbes’ preference for using certain codons to maximize the efficiency of protein production by designing a poliovirus that substituted rarer, less efficient codons in sequences encoding the viral shell. The resulting virus was able to copy itself, albeit more slowly. After the investigators administered the engineered virus to mice, the animals were protected from infection when they were later exposed to wild poliovirus. This technique for taking advantage of codon bias to create a live but weakened vaccine could be applied to other pathogens as well to produce potent but safer vaccines.
Efficient Genes, Effective Medicine

One lesson scientists can take from the recent discoveries about the effects of silent mutations is to be careful in our assumptions. Confidence that synonymous mutations must be "silent" was widespread when there was no mechanism to connect a silent change with an alteration in protein production. But in light of the striking examples described above, this position is no longer tenable.

Recognizing the power of not so silent mutations is beginning to help investigators improve methods for genetic engineering. Knowing which nucleotides in a gene need to be retained and which could potentially be replaced has an immediate application in biotechnology. Both gene therapy and the industrial manufacture of proteins (such as therapeutic drugs), using animals or microbes, rely on the ability to design and fabricate a gene and insert it into a cell's genome. Creating genes that work efficiently is fraught with difficulties, among them ensuring that the newly introduced transgene is activated by the cell, so that adequate amounts of its encoded protein are produced. This is where sensitivity to the effects of synonymous, but not silent, mutations comes in.

In human genes, most introns seem to be dispensable (only one, usually the first intron, appears to be required for the gene to give rise to a protein). This observation means that transgenes can be made compact by removing introns. It also implies that some silent mutation sites could be tweaked without detrimental effect, because leaving out introns does away with the need for splicing enhancers. Released from that constraint, geneticists could exploit those silent sites for other purposes.

A recent experiment at the International Institute of Molecular and Cell Biology in Warsaw illustrates how silent mutation sites could be manipulated for human benefit. Grzegorz Kudla and his colleagues took three genes and did nothing more than change the relative proportion of specific nucleotide bases at the silent sites, then transferred the altered genes to mammalian cells. Remarkably, the investigators found that increasing G and C content led to gene activity and protein manufacture that was up to 100-fold more efficient as compared with GC-poor versions of the same genes.

The new understanding should also inform efforts to understand the root causes of disease. Central to many hunts for the genes underlying diseases are ongoing genome mapping projects to catalogue genetic variation among humans. By identifying all the point mutations, or single-nucleotide polymorphisms (SNPs), in individuals with a given disease, scientists can now home in on regions of the genome containing gene variants that may cause the disease.

Until recently, such searches assumed that if several mutations in a gene correlate with the presence of a disease, those that change the protein's amino acid sequence must be the causal variants. Indeed, the COMT gene associated with pain tolerance is known to have a mutation that changes one amino acid to another, and that variant was long assumed to be the only cause of intolerance to pain. Yet individuals with very...
Scientists showed that the nonsynonymous nucleotide change and the second synonymous change produce folded mRNA shapes dramatically different from the typical sequence. The resulting alternative mRNA structures caused a 25-fold difference in levels of the COMT enzyme in the cells of low- and high-sensitivity subjects.

It is likely that the causes of some diseases have been wrongly attributed to mutations that change proteins, when in fact synonymous mutations are at fault. Investigators need to keep this possibility in mind as they look for a damaging mutational needle in the genomic haystack. And who knows what additional mechanisms of disease such surveys will turn up?

Do the findings so far mean that silent mutations cause disease frequently? Perhaps. One could argue that silent changes are intellectually interesting and potentially of practical importance, but they are not obviously a cause of much illness. Recent examinations of how genes evolve suggest, however, that such a view may be too complacent. Several years ago one of us (Hurst) showed that in one segment of the BRCA1 gene (associated with early-onset breast cancer), silent sites evolve very slowly in both rodents and humans. In contrast, the other sequences in this gene segment evolve at a normal rate. The difference does not mean that mutations in the slow-evolving silent sites were rare but rather that individuals who carried the mutations died without passing them on. The segment was later shown to coincide with the location of a splicing enhancer—in other words, it was another example of an area where silent mutation was so damaging its carriers died out.

Just how common are gene segments in which natural selection exerts pressure for the silent sites not to change? Hurst and Parmley investigated that question by scanning genes, searching for regions in which the rate of evolution at silent sites differed markedly from the rates at sites that change amino acids. To their surprise, they discovered that DNA segments containing unusually slow-evolving silent sites are relatively common. Indeed, they are more prevalent than regions where encoded amino acids evolve exceptionally quickly. A stretch of very highly conserved silent mutation sites occurs, on average, once every 10,000 to 15,000 nucleotides of gene sequence.

We estimate that between 5 and 10 percent of human genes contain at least one region where silent mutations could be harmful. Peter Schattner and Mark Diekhans of the University of California, Santa Cruz, performed a similar analysis, looking for large regions in genes with unusually hyperconserved silent sites. They estimated that there were about 1,600 such blocks in the nearly 12,000 genes they examined, which approximately corresponds with our estimates. Both these estimates are, however, likely to be low, and the true figure may well be considerably greater. If these conserved sites indicate the locations of silent mutations that can cause disease, as seems probable, ignoring them will inevitably lead to misidentification of disease-associated mutations.

Recognition that natural selection does take notice of not so silent mutations has gone hand in hand with scientists’ realization that the process through which genes make proteins is vastly more complex and more nuanced than previously imagined. The way genes evolve and the ways in which they work are also intimately coupled to a degree barely understood even a decade ago. Further study of both these processes will continue to illuminate the remarkable complexity of the workings of genomes. DNA, for example, is not simply a linear molecule but is wrapped into coils and must be unpacked to enable transcription. Does the control of this process leave a fingerprint on silent sites as well? Likewise, runs of rare codons are more common than they should be, but what are they doing and why?

Answers to these and related questions will not just make greater sense of making proteins, they could well provide insights that cure disease.