Evolutionary Anthropology 15:83–87 (2006)



# Dark Matter, Coming to Light

KEN WEISS

When matter meets antimatter in space, destructive annihilation follows. In life, there is a world of genetic antimatter. But when that meets its match, the annihilation is creative.

In 1887 at Case Western University, Albert Michelson and Edward Morley did an experiment to test the theory that space was filled with a substance called ether, the medium that transmitted light. They compared the speed of light as it traveled in two directions at right angles to each other. If ether existed, then in one direction the earth would be hurtling along with or against the flow of ether, while at right angles the earth would travel across the flow. The speed of light should vary between these two axes just as a train whistle sounds higher when the train is approaching you and lower when it is going away. But Michelson and Morley found no evidence of ether. The apparent emptiness of space led Albert Einstein to his theory of relativity, in which light always travels at the same speed.

Since then cosmologists have had a field day with strange and, yes, ethereal concepts of nature. One is that empty space really isn't empty after all, but is filled with "dark matter" (Fig. 1), invisible because it can't interact with light, although its existence can be inferred from its gravitational effects on the starry world that we can see. There may be 400 times as many dark-matter particles as there are particles of regular matter. A second intriguing concept is that of antimatter. Antimatter is a mirror image of the normal matter we deal with every day but exists, we might say, in a different dimension. Should a body meet an anti-body comin' through the nebulae, their kiss will annihilate each other in a puff of energy (Fig. 1).

Life involves surprisingly similar concepts. But first we need a bit of background. Mendel's experiments showed that the traits of organisms are controlled by inherited particles, later called genes and eventually identified as DNA molecules. Life is based on the sequence nature of DNA or, specifically, the order of its four nucleotide building blocks, A, C, G, and T. DNA sequence is copied, or transcribed, into a single-stranded RNA molecule called messenger RNA (mRNA), which is then translated into a corresponding amino acid sequence of proteins. Proteins are important functional units of life. Transcription works because of complementary base-pairing in which As always pair with Ts, and Cs with Gs. For example, transcription enzymes will move along DNA matching a sequence TGGCGT to make an RNA string with a sequence of ACCGCA. Changes in DNA sequence can change the corresponding amino acids, altering protein function. Generation by generation, by luck and natural selection, that variation wends its way through time in the process we call evolution.

The human genome contains about 25,000 genes; each cell contains copies of all of them. You are an organism rather than a uniform gelatinous blob because your cells formed different tissues by using subsets of those genes during embryogenesis. Developmental biology has become a search for those tissue-specific subsets.

How genetically complex is development and how hard will it be to find? There are two basic scenarios. A complex trait might be controlled by one or a few genes, much as, for Mendel, a single gene controlled smooth or wrinkled peas. He studied particular traits in specific strains of domesticated peas on the farm, but if his findings also represent how life is in the wild it means we can expect that traits we care about, like having five fingers or upright posture, are also determined by one or a few genes. If that is true, we have powerful ways to find those genes. But Mendel's findings seemed to conflict with Darwin's notion that life evolved by slow gradual change.[1] This is because most mutations known at the time caused large and usually disastrous changes, not benign variations like smooth or wrinkled. Yet for most traits in nature, most individuals aren't far from average. Moreover, the classical Darwinian view is that selection tinkers gradually with this basically normal variation. Mendel's discretely varying traits seem irrelevant to this process.

A famous paper by Fisher in 1918 reconciled these two points of view.[2] He showed that if many individual Mendelian genes contribute individually small effects, their aggregate produces the kind of smooth variation we see in traits like the distributions of weight, stature, tooth size, or blood pressure. Fortunately, closer inspection showed that most mutations have only small effects after all, compatible with Darwinian notions of evolution. We had a theory! Unfortunately, that theory has two unsettling implica-

Ken Weiss is Evan Pugh Professor of Anthropology and Genetics at Penn State University. E-mail: Kenweiss@psu.edu.


DOI 10.1002/evan.20104
Published online in Wiley InterScience (www.interscience.wiley.com).

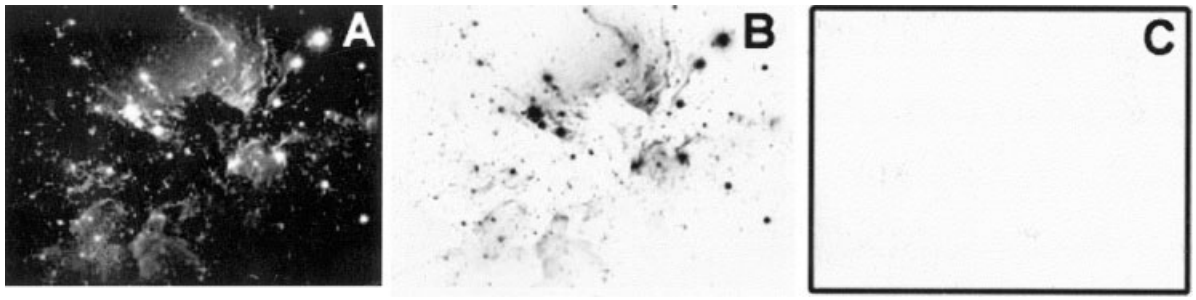


Figure 1. Dark Matter. A. Starfield as we see it. B. Starfield as anti-people see it; C. Union!

tions. First, if complex traits have a multitude of contributing genes of tiny individual effect it will be very hard to find them. Further, the theory implies that many different combinations of existing variation will lead to the same trait. There will not be a single genetic basis for a complex trait like diabetes, oblong crania, or language ability, either within a population or among species.

What we know about interestingly complex traits suggests that they have a mix of these two kinds of genetic causation. There are usually a few genes in which variation has large effects, like greatly raising the risk of a given disease or destroying some focal aspects of language (or making smooth peas wrinkled). These are the highly publicized genes that are easy to find and give a simple view of evolution or raise hopes for genetic cures. However, these major genetic effects account for only a small fraction of the familial variation in the trait; that is, of the correlation of trait values among relatives that appears to be controlled by genes as seen, for example, in twin studies or with measures like heritability. These statements are also true of Mendel's traits in wild peas.[3] The search for the genes of individually small but combined major effect has been a frustrating grope through the dark recesses of our genome. Many remain hopeful that rapidly increasing computing and DNA sequencing technology will eventually shed light on even these dusky genes. But what new technology has been doing instead is to bring discoveries to light that may keep us in darkness. And this brings us back to the strange notions of physics.

## THE DARK MATTER OF LIFE

Protein-coding genes represent only about 5% of our genome. If most of the remaining DNA has any function, we have not been able to see it. However, it now seems that much, even a majority, of our genome is transcribed into forms of RNA other than mRNA.[4,5] Cells contain a huge pool of noncoding RNA of unknown use. What was formerly dismissed as "junk" DNA is now being referred to, more respectfully, as "dark matter." What is it doing?

A natural response is to say that we know how genes work, so RNA dark matter must just be the result of sloppy DNA copying that, like a teenager's room, is just part of the chaos of life. Like dark matter in space, it doesn't interfere with the real thing. However, we've recently learned that some of this RNA is part of elaborate cellular machinery that has been evolving over a substantial portion of the history of life. This is the fraction of the total dark matter that is known as antisense-RNA (asRNA).[6] It has that name because its sequence is complementary to the mRNA transcribed from genes (the "sense" sequence): Where mRNA has C, its corresponding asRNA has G, and so on. Present data suggest that asRNA corresponding to about 10% of all human genes, or twice the genomic fraction of "visible" genes, is transcribed. The fact that this is achieved in various unrelated ways is a hint that something biologically important is going on here. It's also worth taking seriously because, should a mRNA body meet an asRNA anti-body comin' through the cell, their kiss binds them together, with implications I will explain.

Many different kinds of cells have asRNA. Sometimes a gene's mRNA and its corresponding evil asRNA Moriarty are both expressed at similar levels, while under other conditions the relative levels of the two may change in inversely proportional ways, when one is common the other is rare.[4,7,8] Or a cell may contain only one of them at a given time. The patterns are replicable and can't just be sloppiness, because a bit of DNA is usually transcribed only when it is flanked by additional short sequence elements that the transcription machinery (various proteins) uses to grab the DNA there and start moving along making an RNA copy, nucleotide by nucleotide. Too many genes have transcription signals on both strands for them to be there by chance alone. In fact, asRNA for some genes does not come from the gene's second strand, but from DNA somewhere else in the genome, yet both have evolved to be transcribed in the same cells.

Evolution can be devilishly devious. Genomes are littered with repeated short sequences of various kinds. One of these is called a LINE transposable element. LINEs have the occasional ability to make duplicates of themselves that are then inserted elsewhere in the genome. Over the eons, some of these elements have landed in DNA that codes for the tail end of mRNA of various genes (called the 3′ untranslated region of mRNA, for readers who care). This part of the message is not translated into amino acid sequence, so the inserted LINE element doesn't interfere with normal business. We used to consider these inserted elements to be harmless genetic litter, but LINE sequences have also occasionally landed in introns of

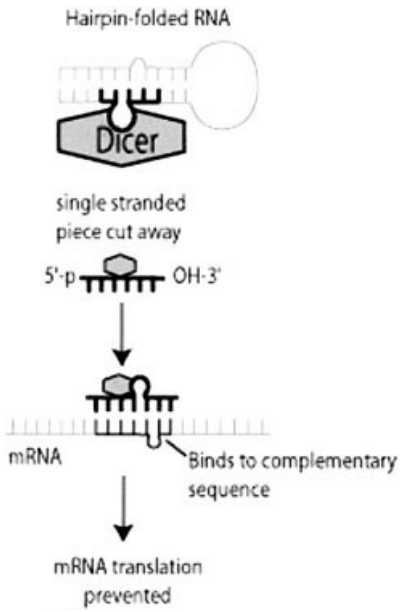


Figure 2. One model of RNA interference. In the nucleus, RNA is transcribed from DNA and folds into a hairpin shape by complementary base-pair binding. Then, in the cell this is captured by a molecule of the Dicer protein, which cleaves out a short RNA sequence. Together, this binds to complementary sequence in an mRNA molecule. (Modified from Weiss and Buchanan.[9])

genes. Introns are noncoding bits of sequence that interrupt the coding sequence of most genes. Like cowbirds laying eggs in other birds' nests, laying a LINE in an intron tricks the cell into transcribing the LINE as part of the immature mRNA for that gene. The LINE is spliced out as the exons are joined to form the mature, active mRNA, but the spliced-out LINE becomes an asRNA having a sequence that is complementary to LINE sequence bits in the tails of other genes' mRNA. That's devilish!

That this kind of asRNA has not been thrown out by natural selection is a further clue that it has important function. But another way asRNA is produced shows unambiguously that it is functional (Fig. 2). From many parts of the genome, short RNA sequences are transcribed, then fold up into a hairpin shape. (To fold up like that, each such RNA evolved to contain inverted repeat sequences that can bind to each other in a complementary way.) That shape is later recognized by a protein called Dicer, which cleaves such hairpins so as to release a small (20–30 nucleotide) bit of the RNA that has a sequence complementary to the tail end of mRNA from one or more genes.[10,11]

The Dicer mechanism is elaborate and phylogenetically deep, found in both animals and plants. It has been lurking quietly in the darkness beneath our century-long theoretical understanding of what genes are all about. But how could something like this ever evolve? Haven't we known for decades that "real" genes code for proteins and that's the way life works?

## AN ANTI-EMBRYO WITHIN US?

No matter what mechanism produced it, when asRNA meets its complementary sequence in mRNA the complementary bits stick together to form a double-stranded molecule. Cells actively search for and destroy double-stranded RNA, but even if they didn't, such RNA can't be translated into protein because it chokes the ribosomes where translation happens. This means that when sense and antisense RNA meet, like matter and antimatter, they cancel each other out. Thus, asRNA is a counterbalance to the previously known mechanisms that cause a gene to be expressed.

It has been found that asRNA works in various ways.[6] In addition to chopping hairpin RNAs, the Dicer mechanism affects gene expression by modifying chromosome packaging and use in the nucleus. These functions may have originally evolved to provide cellular defenses against incoming viruses or to keep transposable elements from uncontrolled proliferation that could be destructive to the genome.[10] It's easy to see how such a defense could have been favored by selection. Subsequently, the antisense mechanisms were adapted to downregulate genes by inactivating their mRNA after it has been produced. This is important in the dynamics of development and homeostasis.

During embryogenesis, timing is everything. Different tissues and organs begin from a very few primordial cells that become different from their surrounding cells. Examples are the small placodes or organizing centers in early embryos that will form the limbs, ears, nose, eyes, or subdivisions of the brain. Some of the genes that prepare these cells for their subsequent jobs are known. It is thought that the timing and concentrations of the proteins they code for need to be under exquisite control, so that differentiation commitments don't go awry. Thus, one function of sense-antisense annihilation may be to dispose quickly of mRNA for genes that are only briefly needed.

There are many asRNAs. It may be that each cell type expresses a large dark-matter combination of them that is comparable in complexity to the mRNAs that are the generators of that cell type. The anti-sense set would inhibit and dampen gene expression on a complex scale (Fig. 3).[13] This is like an embryo having its anti-twin around, or at least each cell having a substantial antisense combination to keep it from getting a swollen head

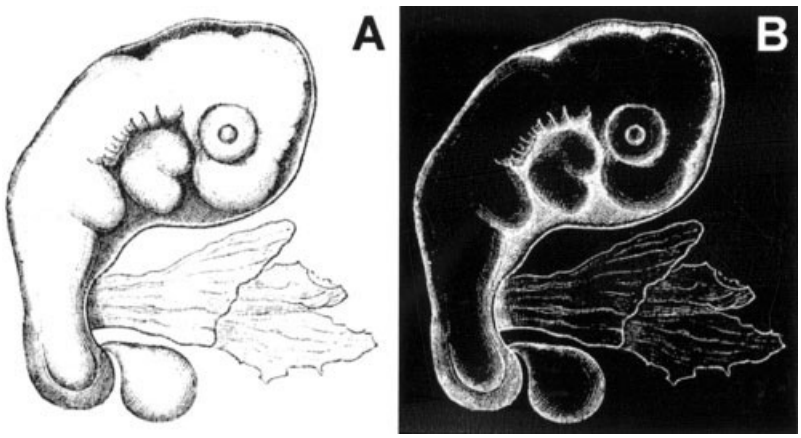


Figure 3. A. An early embryo. B. Its evil anti-twin? Image from the early study of human embryology by William His, taken from Hopwood.[11]

and ensure that it makes a stomach, teeth, and kidneys only where they belong.

Indeed, a new study has found more than 100 types of Dicer-activated asRNA in mouse skin. The expression patterns differ between epidermal and hair follicle cells,[15] showing that there is some sort of genome-wide asRNA regulatory system. When the authors experimentally inactivated the Dicer gene in skin, preventing the asRNAs from being activated, hair follicles evaginated, forming cysts, rather than invaginating to form hair. Individual cells were more likely to die and the mice were not viable. The antiworld appears!

Teeth and hair begin with very similar gene expression patterns as they initiate from small primordial cells. So this study is almost certainly relevant to teeth (we're checking these mice for such effects). There is other direct evidence of asRNA in tooth development as well. The *Msx1* gene is important in the development of vertebrate structures from head to toe, including teeth.[16] If you experimentally interfere with *Msx1* expression in a mouse, you get craniofacial effects like cleft palate and anomalous numbers of teeth. Figure 4 is a schematic induction/repression diagram that indicates how regions of *Msx1* and as*Msx1* expression affect the expression of other genes in patterning the early embryonic jaw. Later in development, the pattern of mRNA and asRNA for *Msx1*, as well as for *Dlx2* and other genes, changes in the two tissues that form the dentine and enamel of each tooth.[17]

These interlocking worlds and antiworlds may help account for phenotypic stability and "canalization" in evolution. Related species may differ mainly by small changes in developmental timing that could evolve rapidly, a suggestion made 30 years ago to account for the seemingly large human-chimp differences that have evolved in only a few million years.[14] Such changes could affect the size or shape or the patterning of modular structures like cranial bones, limbs, or teeth, the stock in trade of anthropology.

Over the years, my lab, like others, has found sequences complementary to genes we were interested in. We dismissed such findings as artifacts because RNA interference wasn't then known. One sees what one is ready to see.

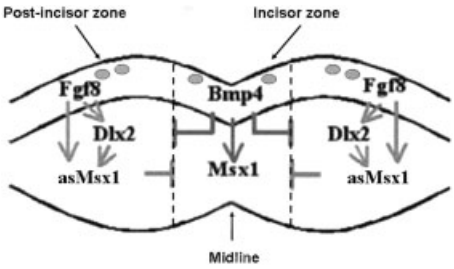


Figure 4. A model for sense-antisense inducing (big arrows) or inhibiting (T-lines) expression of various developmental genes in jaw and tooth development. An embryonic jaw viewed from the front where the two developing jaws join at the midline. There, *Bmp4* expressed in overlying ectoderm induces *Msx1* expression in underlying tissue. Meanwhile, Fgf8 in the ectoderm induces *Dlx2* and antisense-*Msx1* in the underlying tissue in lateral areas. *Msx1* and as-*Msx1* inhibit each other, keeping them zonally expressed. Different tooth types (dark circles) develop in each zone. Modified from Coudert and coworkers.[17]

## IMPLICATIONS FOR COMPLEX LIFE AND EVOLUTION

Evolution still presents surprises. RNA dark matter doesn't change the basic principles of evolution, but does show the limitations of trying to understand evolution strictly by looking at protein-coding genes as we've been doing for nearly a century. asRNA makes the genetic basis of biological traits more complicated, causation more layered, variation and the target for mutation greater, and the potential impact of selection more nuanced. And as I mentioned earlier, many additional kinds of RNA dark matter will surely complicate matters when their functions are known.

This adds to the heterogeneity of causation that Fisher, as early as 1918,[2] showed to be, in a sense, innumerably complex. RNA dark matter and its many manifestations are likely to make traits even less dissectible into individual genetic causes than we had thought. This is a profound lesson, perhaps even a warning. Each new layer of causation makes genotypes less directly primary in biology. It isn't that genes are unimportant or uninvolved, but that the connection is quite indirect, so the phenotype in some ways has a different level of reality. That's because if the same phenotype can be produced by many different combinations of factors the phenotype becomes correspondingly more important relative to the individual genes that affect it as the causal units of life. Even a trait that evolves by consistent, strong natural selection may involve so many contributing genes and anti-genes, with variable timing and intensity, that each instance of a phenotype has a somewhat different genetic basis. For example, the basic structure of teeth has been adaptively maintained for most of vertebrate history, but it appears that different genes, evolving independently underneath this conserved basic structure, are responsible for the mineralized layers of teeth in teleost fish and in we tetrapods who eat them.[18,19]

The implications go further and involve the nature of homology. One objective of the use of model systems like the laboratory mouse is to understand a genetic mechanism that we can extrapolate to other species we can't study, like our fossil ancestors. We have recently been comforted by the discovery of deep phylogenetic conservation of basic developmental genetic mechanisms. But a mechanism shared among widely divergent species can't explain their differences, and as new discoveries continue to add additional layers of causal complexity, our ability to extrapolate with any specificity correspondingly weakens. That doesn't mean we can't understand how complex traits are produced in a general sense. However, it may mean that we can't specify with any precision which of the genes' variations is responsible for the enlargement of Paranthropine molars, the retraction of the face in the evolution of *Homo erectus,* or the turn of the hominid ankle.

Under the old central dogma of biology, gene-makes-protein, we saw no need for antisense mechanisms and, to my knowledge, nobody suspected them. At that time we couldn't see the dark matter of life. Adding numerous levels of control of traits would have

seemed to be a needless burden that evolution would have rejected. But now that we've stumbled onto RNA dark matter and can detect it systematically, it's easy to see in retrospect that since DNA is made of two complementary strands, both strands might be transcribed and various complementary base-pairing dynamics could evolve. Sense and antisense are life's version of Newton's law of equal and opposite reactions.

RNA dark matter acts creatively in sculpting an embryo or responding to environmental changes. asRNA is already used to alter normal development experimentally, and is widely touted as the next miracle gene therapy. It's also a lesson in humility for any of us who confidently called noncoding DNA "junk."

Like dark matter in space, matter and antimatter in life are complementary. When matter and antimatter particles meet in space, they annihilate each other, but produce a burst of energy. In life, when antisense and sense RNA meet, it's a complement that leads to—nothing. In the search for biological causation, that nothing was hard to find when we were looking for something. But just because we can't see something doesn't mean it isn't there. We only know when we've turned on the light. The problem in science is that until we've found it, usually by chance, we're not even looking for a switch.

## NOTES

I welcome comments on this column: kenweiss@psu.edu. I have a feedback and supplemental material page at http://www.anthro.psu.edu/weiss_lab/index.html. I thank Anne Buchanan, Sam Sholtis, Abby Bigham, and John Fleagle for critically reading this manuscript.

## REFERENCES

**1** Weiss KM. 2004. "The smallest grain in the balance." Evol Anthropol 13:122–126.

**2** Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans Roy Soc Edinb. 52:399–433.

**3** Weiss KM. 2002. Goings on in Mendel's garden. Evol Anthropol 11:40–44.

**4** Johnson JM, Edwards S, Shoemaker D, Schadt EE. 2005. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet 21: 93–102.

**5** Mattick JS. 2005. The functional genomics of noncoding RNA. Science 309:1527–1528.

**6** Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. 2004. In search of antisense. Trends Biochem Sci 29:88–94.

**7** Chen J, Sun M, Hurst LD, Carmichael GG, Rowley JD. 2005. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. Trends Genet 21:326–329.

**8** Kiyosawa H, Mise N, Iwase S, Hayashizaki Y, Abe K. 2005. Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. Genome Res 15:463–474.

**9** Weiss KM, Buchanan AV. 2004. Genetics and the logic of evolution. New York: Wiley-Liss.

**10** Almeida R, Allshire RC. 2005. RNA silencing and genome regulation. Trends Cell Biol 15:251–258.

**11** Huttenhofer A, Schattner P, Polacek N. 2005. Non-coding RNAs: hope or hype? Trends Genet 21:289–297.

**12** Hopwood N. 2000. Producing development: the anatomy of human embryos and the norms of William His. Bull Hist Med 74:29–79.

**13** Bartel DP, Chen CZ. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet 5:396–400.

**14** King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188: 107–116.

**15** Yi R, O'Carroll D, Pasolli HA, Zhang Z, Dietrich FS, Tarakhovsky A, Fuchs E. 2006. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. Nat Genet 38:356–362.

**16** Berdal A, Lezot F, Pibouin L, Hotton D, Ghoul-Mazgar S, Teillaud C, Robert B, MacDougall M, Blin C. 2002. Msx1 homeogene antisense mRNA in mouse dental and bone cells. Connect Tissue Res 43:148–152.

**17** Coudert AE, Pibouin L, Vi-Fane B, Thomas BL, Macdougall M, Choudhury A, Robert B, Sharpe PT, Berdal A, Lezot F. 2005. Expression and regulation of the Msx1 natural antisense transcript during development. Nucleic Acids Res 33:5208–5218.

**18** Kawasaki K, Weiss KM. 2005. Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. J Exp Zoolog B Mol Dev Evol. 306B:295–316.

**19** Kawasaki K, Suzuki T, Weiss KM. 2005. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. Proc Natl Acad Sci USA 102:18063–18068.