

Comparative Genomics and Evolution of Alternative Splicing: The Pessimists' Science

Irena I. Artamonova^{†,‡} and Mikhail S. Gelfand^{*,§,⊥}

Group of Bioinformatics, Vavilov Institute of General Genetics, RAS, Gubkina 3, Moscow 119991, Russia, Institute of Bioinformatics, GSF-National Research Center for Environment and Health, Ingostaedter Landstr.1, Neuherberg 85764, Germany, Institute for Information Transmission Problems, RAS, Bolshoi Karetny pereulok 19, Moscow 127994, Russia, and Faculty of Bioengineering and Bioinformatics, Moscow State University, Vorobievsky Gory 1-73, Moscow 119992, Russia

Received October 27, 2006

Contents

1. Alternative Splicing, Its Prevalence and Function	3407
1.1. Popularity of Alternative Splicing	3407
1.2. Prevalence of Alternative Splicing	3407
1.3. Alternative Splicing and Gene Expression	3409
1.4. Alternative Splicing and Proteome	3409
1.5. Other Reviews	3409
2. Evolution of Exon–Intron Structure	3410
2.1. Theories about Intron Origin	3410
2.2. Intron Gain and Loss	3410
2.3. Mechanisms of Intron Insertion	3410
2.4. Exon–Intron Structure and Proteins	3411
2.5. But What about Alternative Splicing?	3411
3. Evolution of the Alternative Exon–Intron Structure and the Functionality of Nonconserved Isoforms	3411
3.1. Low Conservation of Alternative Splicing	3411
3.2. Functionality of Nonconserved Isoforms	3412
3.3. Conserved Exons, Conserved Exons with Nonconserved Splicing, Nonconserved Exons	3412
3.4. Different Exons Behave Differently	3414
3.5. Evolution of Alternative Splicing in Insects and Plants	3414
4. Origin of Alternative Regions	3414
4.1. Sources of New Isoforms	3414
4.2. Duplicated Exons	3415
4.3. Exonization	3415
5. Changes in Alternative Splicing after Gene Duplication	3416
6. Specifics of Selection in Alternatively Spliced Regions and Regulation of Alternative Splicing	3417
6.1. Evolution Rates in Constitutive and Alternative Regions	3417
6.2. Regulatory Sites and Their Evolution	3418
7. Recognition of Alternative Exons and Databases of Alternative Splicing	3419
8. Conclusion	3419
9. Acknowledgments	3422
10. References	3422

1. Alternative Splicing, Its Prevalence and Function

1.1. Popularity of Alternative Splicing

Almost simultaneously with the discovery of splicing,¹ it was recognized that the transcripts of some genes are spliced alternatively, producing several variants of mature mRNA.² The term “alternative splicing” in the current meaning was first used to describe multiple mRNAs produced by the human growth hormone gene.³ Since then, the annual output of papers on alternative splicing has been almost constantly increasing, although the growth has slowed since 1998 (Figure 1a), and the fraction of such papers among all references in the PubMed database reached a maximum in 2002 and began slightly decreasing after that (Figure 1b).

We do not think that the reason for this decline is the general loss of interest. Rather, the field fell a numerical victim of its own success: the mass sequencing of expressed sequence tags (ESTs)^{4–11} and full-length cDNAs^{12–17} had paved way to genome-scale studies, so that alternative splicing of individual genes was not so exciting anymore. On the other hand, it has opened the door to general questions about the emergence and evolution of alternative splicing, and indeed, the fraction of papers in this area is increasing steadily (Figure 1).

1.2. Prevalence of Alternative Splicing

Until the late 1990s, although several important cases of alternative splicing had been described, such as mammalian immunoglobulins¹⁸ and T-cell receptors,¹⁹ the p53 gene,²⁰ various growth factors,^{21–23} the CD44 receptor involved in a variety of lymphocyte functions,²⁴ myofibrillar proteins²⁵ (tropomyosin, myosin, titin, etc.), P-element transposase,²⁶ the sex differentiation pathway of *Drosophila melanogaster*²⁷ and, later, other insects,²⁸ and many others, alternative splicing was considered to be a relatively rare phenomenon occurring, for instance, in about 5% of human genes.²⁹ The true prevalence of alternative splicing was appreciated only after the EST sequencing projects started producing large amounts of mRNA fragment sequences. Mapping of these ESTs to known genomic or mRNA sequences demonstrated that at least one-third of human genes are alternatively spliced.^{30,31} With the arrival of more and more EST data and improvement of bioinformatics algorithms, the estimated fraction of alternatively spliced human genes gradually increased, reaching two-thirds for most human chromosomes (Figure 2).

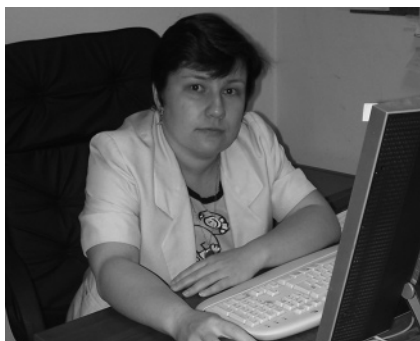
* Corresponding author. E-mail: gelfand@iitp.ru.

† Vavilov Institute of General Genetics.

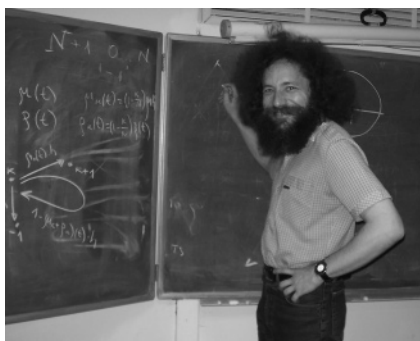
‡ GSF-National Research Center for Environment and Health.

§ Institute for Information Transmission Problems.

⊥ Moscow State University.



Irena I. Artamonova is a group leader in the Vavilov Institute of General Genetics of the Russian Academy of Science in Moscow. She received M.S. degrees (both magna cum laude) from the Lomonosov Moscow State University (pure mathematics) and the Moscow Institute of Physics and Technology (physical and chemical biology and biotechnology) and then moved to bioinformatics working in the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, where she received her Ph.D. degree in molecular biology in 2002. After working as a postdoctoral researcher in the Munich Information Center for Protein Sequences, she won a grant from the RAS program "Molecular and Cellular Biology" and returned to Moscow to start her own group. Her research interests are in comparative genomics, genome structure, gene and genome evolution, alternative splicing, regulation of gene expression, gene annotation, and text and data mining.



Mikhail S. Gelfand is the Head of the Research and Training Center in Bioinformatics and the Deputy Director for Science of the Kharkevich Institute for Information Transmission Problems, RAS in Moscow, Russia, and a Professor at the Department of Bioengineering and Bioinformatics of the Lomonosov Moscow State University. He graduated from the Department of Mathematics of the Moscow State University, received his Ph.D. degree in mathematics from the Institute of Theoretical and Experimental Biophysics, RAS (Pushchino), and the Doctor of Sciences degree in molecular biology from the State Research Institute for Genetics and Selection of Industrial Microorganisms (Moscow). He is a member of the editorial boards of several journals, in particular, *PLoS Biology*, *Bioinformatics*, *BMC Bioinformatics*, *Journal of Bioinformatics and Computational Biology*, and *Journal of Computational Biology*. He received the "Best Scientist of the Russian Academy of Sciences" award (2004), and the A.A.Baev Prize for work in bioinformatics and genomics from the Presidium of the Russian Academy of Sciences (2007). His research interests include comparative genomics, metabolic reconstruction and modeling, evolution of metabolic pathways and regulatory systems, function and evolution of alternative splicing, functional annotation of genes, proteins, and regulatory sites, and metagenomics.

A projection of estimates obtained using smaller EST data sets available for other animal genomes produced comparable numbers, with the plant estimates being slightly lower.³² The number of isoforms per gene was estimated to be higher in mammals than in invertebrates, fruit fly *Drosophila melanogaster*, and nematode *Caenorhabditis elegans*,³³ although the latter observation had been disputed on technical grounds.³⁴ Use of estimates independent of the EST coverage demonstrated a comparable prevalence of alternative splicing

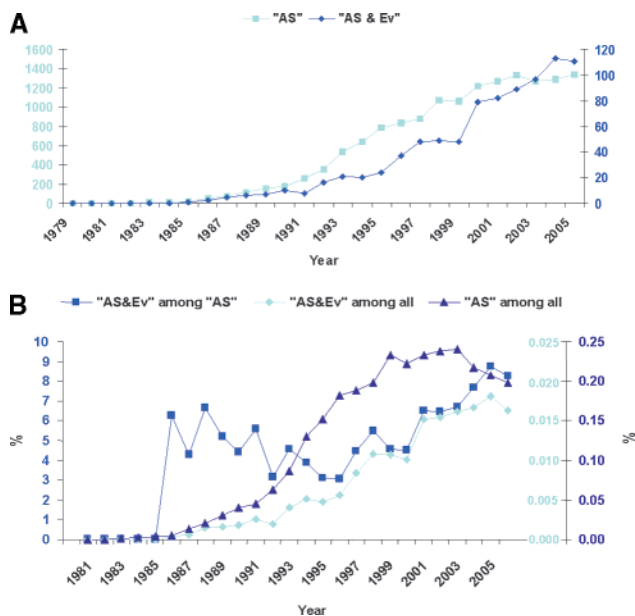


Figure 1. Publications about alternative splicing and its evolution: top, number of abstracts in PubMed with word combination "alternative splicing" (light blue) and "alternative splicing" AND "evolution" (bright blue) per year (horizontal axis); bottom, fractions of abstracts with "alternative splicing" among all abstracts in PubMed (dark blue), with "alternative splicing" AND "evolution" among all abstracts (light blue), or with "alternative splicing" AND "evolution" among abstracts with "alternative splicing" only (bright blue). The colors of scales correspond to the colors of plots.

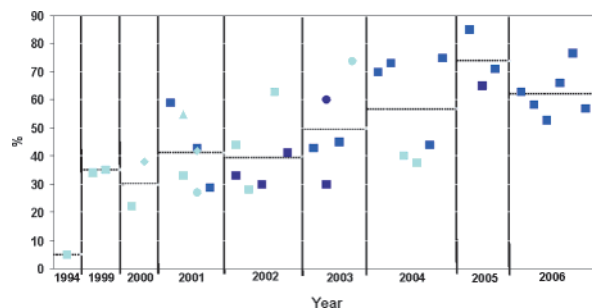


Figure 2. Estimates of the prevalence of alternatively spliced genes among all human genes (light blue), genes from separate human chromosomes (bright blue), and mouse genes (dark blue). The samples consisted of all transcribed genes under consideration (■), genes with high EST coverage (▲), only multiexon genes (●), and mRNA-supported ones among the latter (◆). The averages for a year are shown as level lines. The order of references along the horizontal axis is as follows: 29, 364, 31, 365, 30, 236, 366, 82, 367, 32, 368, 369, 36, 370, 371, 372, 373, 87, 374, 375, 376, 241, 377, 378, 379, 12, 380, 381, 382, 383, 384, 385, 386.

in the human and chicken (about 43%), followed by rodents (about 32%), and then invertebrates *Ciona intestinalis*, *D. melanogaster* and *C. elegans*.³⁵ A problem here is that not all ESTs may arise from functional alternative isoforms, and rare ones may in fact reflect splicing errors³⁶ (see section 3.2 for a detailed discussion).

The estimated fraction of alternatively spliced *Arabidopsis* genes ranges from 10–12% (refs 35 and 37) to 18% (ref 38) to slightly more than 20% (refs 39 and 40), and remains consistently lower than that of animal genes, with a considerably larger fraction of retained introns^{35,41} (30% of all alternative splicing events compared with less than 10% in animals). These estimates were confirmed not only by EST data but also by the whole-genome microarray analysis.⁴⁰

Very similar estimates were obtained for rice,³⁸ but in other plants the situation might be different, as alternative splicing seems to be more prevalent in most of plant species with sufficiently large EST collections.⁴² At that, *Lactuca sativa* (lettuce, an eudicot plant like *Arabidopsis*) and *Sorghum bicolor* (a monocot plant like rice) demonstrate a more than 3-fold higher rate of alternative splicing.

Among unicellular eukaryotes, alternative splicing seems to be relatively rare: only three genes of the budding yeast *Saccharomyces cerevisiae* were reported to be alternatively spliced;^{43,44} alternative splicing was observed also in a related hemiascomycete *Yarrowia lipolytica*,⁴⁵ in a more distant fission yeast *Schizosaccharomyces pombe*,⁴⁶ and in the malaria parasite *Plasmodium falciparum*.⁴⁷

1.3. Alternative Splicing and Gene Expression

In any case, it is clear that in multicellular eukaryotes, alternative splicing is one of the major mechanisms for generating protein diversity.^{48,49} It is closely linked to other types of events that lead to expression of different mRNAs from one gene, in particular, initiation of transcription from alternative promoters^{50–57} and alternative polyadenylation.^{58–60} In an extreme case, two completely different proteins having no common residues may be expressed from one gene; 12 such cases were identified during reannotation of the *D. melanogaster* genome.⁶¹ Alternative mRNA isoforms not only are translated into different proteins but may be subject to different regulation by internal initiation of translation,⁶² mRNA editing,^{63,64} mRNA decay,^{65,66} and microRNA binding,⁶⁷ reviewed in ref 68. Alternative splicing is also linked to the rate of transcription elongation^{69,70} or initiation,⁷¹ reviewed in ref 72, and to transcriptional readthrough, leading to formation of fused proteins encoded by adjacent genes.^{73,74} Furthermore, alternative splicing is one of the major mechanisms for regulating expression of splicing factor genes, thus creating a potential for various regulatory feedback loops,^{75–78} and this mechanism may be conserved at large evolutionary distances, for example, in vertebrates⁶⁶ and in plants.^{79,80}

Alternative splicing is frequently tissue-specific,^{81–85} with the greatest diversity observed in brain, testis, and liver,⁸⁶ and some similarity in the isoform spectra between related tissues.^{87,88} The preferred types of expressed alternatives (e.g., alternative splice sites vs cassette exons) also may depend on the tissue; for example, the abundance of alternative sites in liver is considerably higher than that in other tissues, and this has been linked to the diversity of splicing regulatory proteins expressed there.⁸⁶ Thus, alternative splicing, working independently of transcription regulation,⁸⁸ provides an additional level of flexible control of tissue-specific gene expression.

Numerous studies demonstrated the existence of cancer-specific isoforms, although it is not immediately clear whether they represent regression to a wrong functional state or simply reflect overall deregulation of cellular processes.^{85,89–93} However, in some cases the medical relevance of these isoforms has been demonstrated,^{94–99} reviewed in refs 100–102. Protein fragments that are different between cancer-specific and normal isoforms^{103,104} have been suggested to be possible targets for specific drugs. The medical importance of alternative splicing is not limited to cancer, as the relative expression level of alternative isoforms may have many different functional consequences. Some genetic diseases are caused by mis-splicing¹⁰⁵ or mutations that create

novel isoforms or alter the ratio of isoforms.^{106–108} Alternative splicing may also be important in pharmacogenomics applications.¹⁰⁹

1.4. Alternative Splicing and Proteome

On the protein level, alternative splicing tends to retain the overall protein structure. It avoids disrupting domains and instead tends to shuffle entire domains^{110–112} and transmembrane segments¹¹³ as opposed to a random control, reviewed in ref 114. A recent study using full-length mouse cDNAs demonstrated that among alternatively spliced genes, about 6% encode cytoplasmic and membrane-bound protein isoforms, 4% encode cytoplasmic and secreted isoforms, and 4% encode membrane-bound and secreted isoforms.¹¹⁵ Alternative regions tend to be located on the surface of protein globules¹¹⁶ and affect disordered regions.¹¹⁷ On the other hand, one report suggested that alternative protein isoforms encoded by as much as two-thirds of alternatively spliced genes would have large conformational differences, since alternative splicing involves regions encoding the protein structural core.¹¹⁸ It has been demonstrated that alternative splicing disproportionately frequently involves domains functioning in signal transduction, cell communication, protein–protein interaction, and similar processes.^{118–120}

The latter findings may reflect the somewhat less appreciated role of alternative splicing in the maintenance of protein identity. Indeed, alternative splicing is a perfect mechanism for generating proteins that differ in only one specific region, thus creating an opportunity for various combinations of properties. The most obvious example is provided by the membrane-bound and secreted isoforms of proteins that differ from cytoplasmic isoforms by the presence of alternatively spliced transmembrane helices and signal peptides, respectively.¹²¹ Indeed, regions encoding N-terminal transmembrane helices and signal peptides are subject to frequent alternative splicing.^{113,122} Alternative splicing has a tendency to target regions corresponding to protein functional sites,^{111,123} often producing proteins with dominant-negative regulatory effect, numerous examples of which are listed in ref 121. This tendency is particularly interesting in the case of transcription factors,^{112,124} producing regulators with different DNA-binding specificities^{125,126} or isoforms having overall dominant-negative effect,¹²¹ and in genes encoding protein kinases and phosphatases.^{127,128}

1.5. Other Reviews

Even this brief introduction demonstrates the functional importance of alternative splicing. Several recent reviews concern various aspects of alternative splicing,^{68,101,114,121,129–135} including technical,^{136–138} popular, and philosophical ones.^{139,140} There also exist several databases on alternative splicing listed, for example, in ref 131 or in ref 141 (see also <http://www.oxfordjournals.org/nar/database/subcat/1/3>).

Here we concentrate on the evolution of alternative splicing. This area also has been a subject of several excellent reviews,^{142–145} but it develops very fast, and thus new interesting studies appear monthly, if not weekly. One of the consequences of that is a large number of standing controversies, and one of the purposes of this review is both to list conflicting observations and to suggest possible ways to their resolution and reconciliation.

2. Evolution of Exon–Intron Structure

2.1. Theories about Intron Origin

Prior to detailed discussion of the evolution of alternative splicing, we will briefly consider the evolution of the exon–intron structure itself and the problem of intron origin (for more detailed reviews, see, for example, refs 146–148). The latter has been discussed for about 30 years along two main lines, named “introns early” and “introns late”.

According to the “introns early” theory, introns were present in the common ancestor of eukaryotes and prokaryotes and were subsequently lost in the latter by selection toward compact genomes and short generation time.^{149–151} This theory is based on the observation that in many cases exons coincide with protein domains, either functional or structural, and thus may represent primordial genes encoding early functional peptides.

The “introns late” theory posits that introns were inserted into eukaryotic genomes after their split from prokaryotes.^{152–157} Indeed, some unicellular eukaryotes have no or very few introns, and there exist numerous examples of recent intron gain. While early papers analyzed the exon–intron structure in selected gene families,^{158–167} the most prominent debate probably being the one concerning triose-phosphate isomerases,^{168–170} recent studies take advantage of large-scale genome comparisons. As one could expect, these analyses demonstrate that the extremes of both theories are wrong, as there exist both clearly ancient exons and recently inserted introns.

2.2. Intron Gain and Loss

The main types of events in the evolution of the exon–intron structure are intron loss and gain, intron sliding, exon duplication and other types of exon shuffling, and exonization of intronic regions. The distribution of introns in extant genomes is governed by the balance between the lineage-specific rates of intron loss and gain.^{171–174} Unfortunately, the estimates of these rates crucially depend on the evolutionary model, and for example, very different estimates are obtained when the parsimony¹⁷³ and maximum likelihood with¹⁷² or without¹⁷⁴ parallel gains are used. Estimates of the fraction of intron positions experiencing independent parallel gains in different species also differ^{172,175} but less dramatically. Remarkably, all these estimates were obtained on the same set of 684 genes, introduced in ref 173.

The saturation of the taxonomy tree with sequenced genomes provides for far better resolution. The comparison of human and mouse genes with each other and with rat and dog genomes demonstrated intron loss, mainly in the rodent lineage and mainly of small introns.¹⁷⁶ The loss rates (intron loss per million years per intron) were estimated to be between 4.28×10^{-5} (for human) and 1.8×10^{-5} (for rat). Two more very recent studies of multiple draft vertebrate genomes¹⁷⁷ or fragments of 17 vertebrate genomes homologous to a specifically selected 1% of the human genome¹⁷⁸ (the ENCODE project,^{179,180} see also http://www.nisc.nih.gov/open_page.cgi?path=/projects/encode/index.cgi) considered both constitutive and alternative exons and are addressed in more detail in section 3.1.

Some other examples of such densely populated branches are numerous yeast genomes¹⁸¹ and more than a dozen *Drosophila* genomes¹⁸² (see also <http://www.genome.gov/11008080>). Already, though the average rates are difficult

to estimate, the comparative analysis of multiple genomes makes it possible to discriminate between clearly ancient (predating separation of main eukaryotic lineages) and recent (lineage-specific) introns and exons. This allowed for the comparison of properties of such introns and exons and yielded important observations about the events shaping the exon–intron structure.

The elimination of introns continues in fungi.^{183–186} More detailed analyses show that not only do small, reduced eukaryotic genomes tend to have fewer introns but in such genomes introns tend to occur in the 5′-end region of the gene.^{171,187} This points to the recombination with reverse-transcribed mRNA as a possible mechanism of intron loss,^{188,189} although the evidence is not unequivocal: fungi, despite clear prevalence of 5′-introns, do not demonstrate relatively more frequent 3′-intron loss,¹⁸⁵ and thus the 5′-intron prevalence may be explained by preferred intron gain at the 5′-end.¹⁸⁹ Some 3′-end bias of lost introns was observed in mammals,¹⁷⁶ and while the analysis of recently duplicated paralogs in vertebrates produced no such pattern, all cases of multiple intron losses were consecutive, indicating involvement of spliced cDNA.¹⁹⁰

Analysis of orthologous genes¹⁷³ and paralogous gene families¹⁹¹ demonstrates numerous intron insertions in genes duplicated prior to the plant–animal split. On the other hand, intron losses and especially gains are extremely rare in the recent evolution of plants and mammals,^{191,192} thus demonstrating that the pattern of intron gain and loss is very nonuniform as regards the time scale. However, the above results were obtained by comparison of multiple animal genomes with the single plant genome *Arabidopsis* and thus may not account for the events in the major plant lineages. Indeed, while the exon–intron structure of land plant genes seems to be highly conserved, it was markedly different from the structure of orthologous genes from green algae and other distant plant groups.¹⁹³ Analysis of segmentally duplicated genes in *Arabidopsis*¹⁹⁴ and rice¹⁹⁵ using orthologous genes from other plants as outgroups produced different results: some preference of gains in the former (56 gains vs 39 losses) and clear prevalence of losses in the latter (5 gains vs 49 losses). Still, these results demonstrate a non-negligible rate of intron gain, in contradiction to other studies.

Finally, numerous intron gains and losses were observed in the genome of chordate *Oikopleura dioica*.¹⁹⁶ Frequent intron loss and much rarer gain (up to 10-fold) were found in the genomes of nematodes.^{197–200} Slow rates of loss and possibly no intron gains were observed in the genomes of two *Plasmodium* species.²⁰¹ Finally, about 60% of intron positions in a deeply branching protist *Trichomonas vaginalis* are shared with at least one other eukaryote.²⁰² All these observations seem to indicate intron explosions in early eukaryotes and early multicellular organisms, followed by longer periods of predominant intron loss.^{146,147}

2.3. Mechanisms of Intron Insertion

The mechanisms of intron insertion are not clear, as most attempts to identify homologous introns in non-homologous genes failed;^{203,204} and some similarity observed in nematodes between recently gained introns and other introns¹⁹⁸ was largely limited to minisatellite and palindrome sequences.¹⁴⁷ However, the superficial features of this process were studied in detail. It was suggested that certain sites in exons are prone to intron insertion; these sites were named proto-splice sites.^{159,205} Such sites could contain the exonic parts of the

splicing signals necessary for the successful splicing of the inserted introns.

The frequencies of intron phases (phase 0, introns between codons; phase 1, introns between the first and second codon positions; phase 2, introns between the second and third codon positions) demonstrate prevalence of the phase 0 introns and seem to agree with what could be expected from random intron insertion into proto-splice sites.^{206,207} These studies considered actual intron insertion events, whereas earlier studies that simply considered the reading frame distribution of proto-splice sites produced different results.²⁰⁸

Again, some contradictions were resolved by separate consideration of old and young introns, although this in turn created new problems. Recent introns tend to be phase 0,^{173,198} and on the other hand, such introns are preferentially lost.¹⁸⁸ Preferential intron loss occurs at the 3'-end.^{173,188} Taken together, these observations would mean that phase 0 introns should be relatively rare at the 3'-end, but this is not observed: in fact, the fraction of the phase 0 introns increases toward the 3'-end, especially in high-intron-density genomes.²⁰⁹ Another matter of debate is whether phase 0 introns are predominantly young or old.^{173,210,211} It seems that careful, lineage-specific reanalysis of the data is required to resolve the issue.

Theoretically, what we believe to be proto-splice sites could be the result of convergent evolution caused by the spliceosome requirements on the exonic parts of the splice sites. This possibility was rejected using the analysis of intron insertions into sites that were invariant (on the protein level) in distant lineages.²¹² Additional analysis demonstrated that, while the total information contents of ancient and new splice sites are roughly equal, the exonic parts of new splice sites are relatively stronger compared with old ones.²¹¹ This observation has led to a model in which introns are preferentially inserted into proto-splice sites and then intronic positions of the splice sites mutate toward consensus, freeing the exonic positions to evolve under selection on the protein level. This may explain the observed prevalence of phase 0 introns, since in such introns the most important exonic site position, the last position in the exon immediately adjacent to the downstream intron, coincides with the third codon position and thus in many cases is synonymous. This also is consistent with the idea that the phase 0 introns have a higher probability of fixation^{213,214} and with the observed mutual compensation of substitutions in the exonic and intronic parts of donor splice sites in the human–mouse comparison.²¹⁵ The fact that old introns do not show the expected clear prevalence of phase 0 (although subject of debate, see above) has been explained by intron sliding.²⁰⁴

2.4. Exon–Intron Structure and Proteins

On the protein level, the detailed analysis with new, genome-scale data sets showed that the basic consequence of the “introns early” theory, the correlation between the boundaries of structural protein domains and the intron positions, holds for genes encoding proteins common to prokaryotes and eukaryotes, as opposed to eukaryote-specific ones;²¹⁶ for ancient introns;²¹⁷ and for phase 0 introns and exons containing so-called ancient conserved regions but neither for phase 1 and 2 introns nor for recent exons.^{216,218} Similarly, functional protein domains tend to correspond to exons bounded by same-phase introns that do not interrupt the reading frame^{219,220} although in this case the strongest correlation was found for the 1–1 exons (bounded on both

sides by the phase 1 introns) and not for the 0–0 exons. Similarly, the strongest signal for exon shuffling was observed for the 1–1 exons.²²¹ The apparent contradiction between the preferences for the phase 0 introns and for the 1–1 exons may have consequences for the theory, because the latter observation seems to lack explanation in the frame of the “introns early” theory but can be easily explained by relatively recent, eukaryote-specific exon shuffling.

While the issue of introns in prokaryotes is not unambiguously resolved yet, it is more or less accepted that ancestral or at least very early eukaryotes had spliceosomal introns^{222–228} and that ancestral Metazoa had numerous introns, resembling in this respect extant vertebrates,^{229,230} while many of these introns were lost in the nematode and insect lineages.^{173,191,221,231,232} Moreover, it has been suggested that the burst of exon shuffling²³³ and intron insertion^{173,191} coincided with the emergence of multicellular eukaryotes. A more daring hypothesis links the origin of the eukaryotic nucleus to the necessity for a specialized splicing compartment, which had become required after intron invasion from endosymbiotic mitochondria.^{146,234} To determine whether these early eukaryotes were intron-rich or intron-poor, either an intron-rich, early branching eukaryote needs to be identified, or numerous intron-poor genomes need to be analyzed.

2.5. But What about Alternative Splicing?

Somewhat surprisingly, almost none of these studies explicitly controlled for possible alternative splicing, whereas at least some of the above discrepancies may have been caused by alternatively spliced genes and exons. Indeed, recent studies seem to indicate that in many respects alternatively spliced genes and exons behave as very young ones, and this may exaggerate or obscure the “constitutive” phenomena. On the other hand, a promising theory²³⁵ states that many splicing innovations (sites, exons, isoforms), now constitutive, initially existed as alternative variants.

3. Evolution of the Alternative Exon–Intron Structure and the Functionality of Nonconserved Isoforms

3.1. Low Conservation of Alternative Splicing

The first papers that considered the evolution of alternative splicing on the genomic scale appeared in 2003, when already existing large EST data sets were supplemented by two eukaryotic genomes, human^{236,237} and mouse,²³⁸ that could be used for comparisons. The first attempts to compare alternative splicing patterns in the two species produced a somewhat unexpected result: a large fraction of human alternative regions could not be aligned to the orthologous mouse genes and vice versa.^{235,239} Despite differences in the analysis details and data sets, both studies produced a fairly consistent estimate: about one-fourth of human alternatives are not present in the orthologous mouse genes, and about half of alternatively spliced genes have species-specific isoforms. Even this number could overestimate the prevalence of conserved alternative splicing, because this estimate covered only alternatives that could not be aligned to the mouse genome at all or whose splice sites were destroyed by mutations at the invariant positions, GU (donor sites) or AG (acceptor sites). Thus these studies ignored the possibilities that an exon is spliced alternatively in human but constitutively in mouse or that a splice site is destroyed by

a mutation in an important but not invariant position. On the other hand, as it had been immediately pointed out,²⁴⁰ a large fraction of EST-derived isoforms could represent splicing noise, and thus there were no reasons to expect conservation of such isoforms. Indeed, the fraction of human EST-derived splice junctions with homologs in the known mouse transcriptome turned out to be considerably lower among splice junctions with low EST support: 8% against 61% for predominant splices.³⁶ However, only 10–11% of orthologous splice junctions were present in both transcriptomes,^{36,241} and the attempts to extrapolate this to a hypothetical complete transcriptome produced wide error margins,²⁴² making this type of evidence somewhat inconclusive.

3.2. Functionality of Nonconserved Isoforms

Further debate unfolded in two directions. The proponents of functionality of minor nonconserved isoforms noted the fact that very similar results were observed not only with ESTs but with relatively more reliable mRNAs²³⁹ and that many minor, nonconserved isoforms demonstrated tissue-specific expression patterns and were represented by multiple ESTs from independent clone libraries.²³⁵ The opponents demonstrated that the majority of nonconserved cassette exons are frameshifting (their length is not a multiple of 3) and contain stop codons, and thus the corresponding isoforms encode severely truncated proteins and are candidates for elimination by nonsense-mediated decay.²⁴⁰ It should be noted, however, that isoform function should not be necessarily limited to encoding a protein, as an isoform may also play a regulatory role. Further, young exons, both constitutive and alternative, tended to occur in nontranslated regions.¹⁷⁷

A regulatory role for alternative splicing was established for several genes encoding ribosomal proteins of *Caenorhabditis elegans*²⁴³ and human.⁶⁵ Some of their isoforms contain premature stop codons and thus are subject to nonsense-mediated decay (NMD), a recently discovered surveillance mechanism against transcripts that contain exon–exon junctions downstream of a termination codon.²⁴⁴ This NMD-linked splicing occurs at the excess of the encoded protein that binds to the transcript and inhibits splicing of an alternative intron. Similarly, human splicing factor SC35 autoregulates its expression by producing isoforms subject to NMD.⁷⁸ Other known examples of NMD-coupled regulation are glutaminase²⁴⁵ and fibroblast growth factor receptor 2.²⁴⁶

Indeed, a large-scale computational study demonstrated that as much as one-third of reliable mRNA isoforms (supported by multiple ESTs) may be subject to NMD.²⁴⁷ The authors suggest that RUST (regulated unproductive splicing and translation) is one of the major regulatory mechanisms for establishing the right level of gene expression. A subsequent study, reporting preliminary experimental results, validated some of the potentially NMD-inducing isoforms.²⁴⁸ In another comparative study, the fraction of human cassette exons yielding NMD-inducing isoforms was somewhat lower, about one-fifth, but importantly, a considerable fraction of such exons was conserved in the mouse genome: a third of conserved cassette exons were frameshifting, and moreover, a quarter of conserved isoforms were NMD-inducing.²⁴⁹ Similar results were obtained in a study of all types of elementary alternatives (cassette exons, donor and acceptor sites) in the human, mouse and dog genomes.²⁵⁰ Further, many stop-codon-containing exons were directly demonstrated to be expressed and to cause NMD by

microarray analysis of NMD-inhibited cells.⁶⁶ Interestingly, such exons are extremely highly conserved in mammalian genomes and disproportionately often present in genes encoding splicing factors.

On the other hand, the hypothesis of the widespread regulatory role of NMD-inducing isoforms was challenged in a microarray study that demonstrated that the level of such isoforms is uniformly low in a variety of diverse cells and tissues.²⁵¹ It should be instructive to compare these results with the EST-based observations about tissue specificity of minor isoforms:²³⁵ although formally there is no direct contradiction, since the EST-derived isoforms were not specifically analyzed for the presence of NMD-inducing premature stop codons, indirect reasoning based on other observations makes it rather likely that very often minor isoforms are indeed candidates for NMD.

In plants, that is, *Arabidopsis* and rice, many retained introns are frameshifting.³⁸ However, a large fraction of intron-retaining isoforms were demonstrated to be associated with polyribosomes, thus demonstrating that they are exported from the nucleus and translated,⁴¹ and conserved in the *Arabidopsis*–rice comparisons (see section 3.5) and even in much more distant land plants such as mosses.^{79,80}

3.3. Conserved Exons, Conserved Exons with Nonconserved Splicing, Nonconserved Exons

One important distinction needs to be made at this point. There are two types of aberrant (or, for that matter, bona fide species-specific) cassette exons. One is exons arising from splicing errors produced by the use of cryptic sites. The other type is constant exons that are erroneously (or species-specifically) skipped. The analysis based on the alignment of experimental cassette exons from one organism to the genome of another organism takes into account only the former type of events; its other acknowledged drawback is that it does not allow one to discriminate between errors and bona fide species-specific cassette exons.²⁵² Indeed, even simple triple human–mouse–dog comparisons demonstrate that a considerable fraction of human cassette exons that are lost in mouse are still conserved in dog and vice versa; the level of conservation depends on the inclusion level and frame preservation.²⁵⁰ Similarly, about 60% of exons conserved in mouse and rat are not conserved in human, and about 20% of exons conserved in human and one rodent are not conserved in the other one.²⁵³

Such triple comparisons are sufficient for the estimation of the loss of alternative exons or sites; however, to estimate the rate of cassette exon gain, more genomes are needed, since the exon gain needs to be distinguished from aberrant splicing (Figure 3). A mouse–rat–human comparison with pig ESTs as an outgroup was used to estimate the rate of exon birth in rodents²⁵⁴ (new exons were defined as exons conserved in mouse and rat, and missing in the human genome and pig ESTs). This rate was estimated as 2.7×10^{-3} per gene or 81 per genome per million years in the mouse lineage. As expected, it turned out that 87% of rodent-specific exons were alternatively spliced; however, 48% of them were represented by single ESTs and further 27% by rare (two to five) ESTs. It is not clear to what extent the mouse–rat conservation alone is sufficient to claim the functionality of such low-supported exons. Indeed, mouse exons that could not be aligned to the rat genome were not considered at all, and thus some conservation is expected simply by definition: in a conserved region in DNA, chance

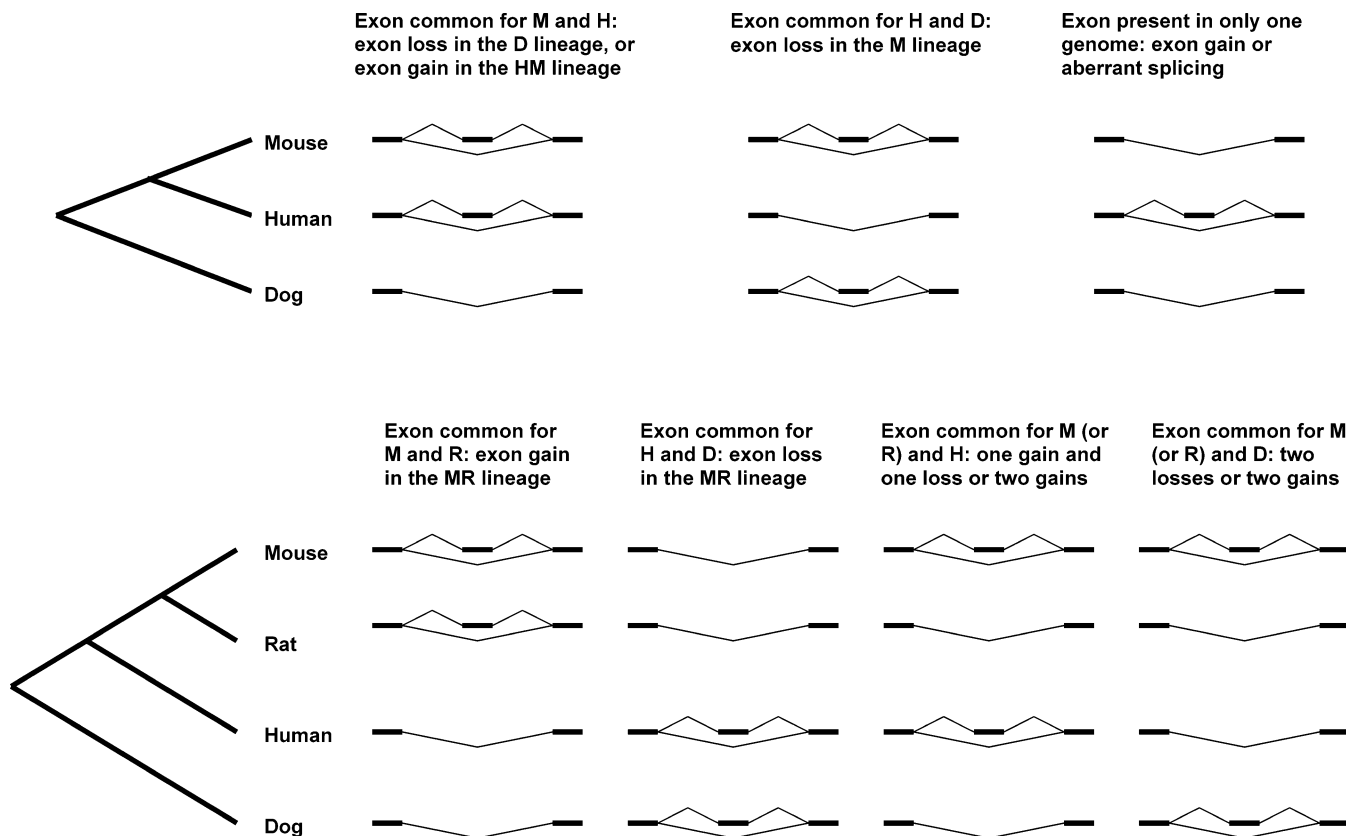


Figure 3. Estimation of the exon loss and gain rates. Three genomes (two sister lineages and an outgroup) are sufficient to estimate the rate of exon losses after the split of sister lineages but not to distinguish exon gains and aberrant splicing. Four genomes allow distinguishing between the latter using conservation in sister lineages as a proxy for functionality.

activation of cryptic sites would create a seemingly conserved exon.

A somewhat conflicting observation was made in another four-genome study (human, mouse, rat, and dog), where intron losses disrupting alternative splicing were not more frequent than expected given the fraction of alternatively spliced introns.¹⁷⁶ However, this study used a different approach: intron gains and losses were considered instead of exon gains and losses, and thus these observations are not directly comparable.

Somewhat more reliable estimates, confirming the initial observations of lower conservation of alternative exons compared with constitutive ones but still leaving room for different interpretations, were obtained in very recent studies that compared multiple vertebrate genomes. In one such study, human exons were mapped to the genomes of seven other vertebrates¹⁷⁷ (human, chimpanzee, mouse, rat, dog, chicken, zebrafish, fugu). This approach has a number of appealing features. First, since exons conserved in at least two genomes were analyzed, they could be assumed to be functional, although conservation in human and chimpanzee does not prove that much (cf. section 3.2). Second, this technique allowed for identification of the time of exon gains, as each successive genome served as an outgroup. More exactly, if, for instance, a human exon had a dog ortholog but lacked a rodent ortholog, it was assumed to be gained by a common ancestor of dog and human (the authors assumed that carnivores are a sister taxon of primates to the exclusion of rodents, contrary to an accepted taxonomy in, say the Human Genome Browser,²⁵⁵ although the latter served as a data source). A problem with this approach is that it does not account for the implicit possibility of exon

losses in the outgroup in each particular comparison, while the possibility of exon loss in ingroups (intermediate taxa, for example, chimpanzee in the above example) is ignored explicitly. Nevertheless, technical problems aside, this study produced results consistent with other studies: the fraction of cassette exons is consistently increasing as relatively young exons are considered.

In a similar study,¹⁷⁸ human genes were compared with 17 vertebrate genomes, again using the Human Genome Browser global alignments.²⁵⁶ If a human exon could not be aligned to the corresponding intron in the orthologous gene, the closest outgroup was used to decide whether it had been exon gain in the human gene or exon loss in the orthologous gene. The results were qualitatively the same as in the above study: the rate of exon loss was roughly constant in time, the rates of constitutive and major isoform cassette exon gains were similar and also constant, whereas the rate of minor isoform exon gain was increasing when genomes closer to human were considered. This could be the evidence that human minor isoform exons were young, but an important control of their functionality was missing. Indeed the same trend could be expected if these exons were aberrant, given that the degree of conservation (alignability) of intergenic regions is decreasing as one moves from chimpanzee to more and more distant vertebrates.

The splicing errors leading to exon skips are important, because there is a logical possibility of conserved (on the sequence level) exons that are constitutively spliced in one species, but may be skipped in another one. The fraction of such exons among all cassette exons has been estimated to exceed 11%.²⁵⁷ In a smaller-scale, but more direct experiment that considered splicing of 20 genes in 11 tissues in human

and mouse, only four genes demonstrated the presence of the same isoforms; of the remaining genes, almost all had exons that were alternative in human but constitutive in mouse⁸⁴ (the apparent asymmetry was caused by the fact that alternatively spliced human genes were initially selected for the study). Nonetheless, such exons may not be found using simple analysis of genomic conservation, as they are conserved in any case. Instead, a detailed experimental analysis of transcriptomes of both species is required. Probably the only species with sufficiently high EST coverage is human, and because of that, simple mapping of human EST-derived alternatives to any other, obviously incomplete transcriptome is not sufficient.^{36,241,242} However, a large-scale application of the microarray technology to the mouse transcriptome and the comparison of these data with the human ESTs has led to delineation of three classes of cassette exons: those that are alternatively spliced both in human and in mouse, mouse-specific ones that do not exist in the human genome, and exons that are alternatively spliced in mouse but are constitutive in human.⁸⁸ It turned out that >70% of mainly skipped exons are genome-specific ones, in agreement with early EST-based observations.²³⁵ In contrast, a majority of rarely skipped exons are constitutively inserted in human. Overall, less than 20% of mouse cassette exons seem to correspond to alternatively spliced exons in human: this is an implausibly low estimate, although still higher than purely EST-based ones. A possible explanation is that, first, the majority of mouse mainly inserted cassette exons may be constitutive ones, skipped as a consequence of mis-splicing, and second, there is always a possibility of some unconsidered cell type, stage, or condition in which a human exon is skipped: in many cases dedicated analysis indeed finds splice junctions conserved in a genome but not in a transcriptome.²⁵⁸ Both types of species-specific alternative exons (corresponding to constitutive exons or completely nonconserved) tend to be potentially NMD-inducing (upon exclusion and inclusion, respectively).²⁵¹ In addition, it should be noted that EST data demonstrated strong correlation between exon inclusion levels in human and mouse.^{235,249,258} Similarly, a PCR-based method was used to establish conservation of developmental variations in the isoform abundance in nematodes *C. elegans* and *C. briggsae*.²⁵⁹

3.4. Different Exons Behave Differently

It seems clear that preferentially included (major form) and preferentially skipped (minor form) cassette exons have different properties. The major form exons tend to be as conserved as constitutive ones²³⁵ and may even correspond to constitutively included exons;⁸⁸ the minor form exons are frequently species-specific.^{88,235} Exons with clear tissue-specific inclusion patterns show a marked preference to be frame-preserving, in comparison to exons with relatively uniform skipping/inclusion ratios in different tissues;²⁶⁰ on the other hand, conserved frameshifting exons tend to be preferentially included.²⁶¹ The major form exons tend to affect protein domains more frequently than the minor form ones.⁸⁸

In the same vein, conserved and species-specific exons also tend to have different functional properties. Conserved minor form exons show a marked tendency for frame preservation, compared with the major form and constitutive exons that have only a weak such preference.²⁶² Conserved frame-preserving cassette exons tend to avoid overlapping with protein domains, whereas frameshifting ones are similar

in that respect to constitutive exons.²⁶¹ Conserved cassette exons are enriched in brain tissues, whereas nonconserved ones are enriched in testis and cell lines.²⁵⁸ Exons with nonconserved alternative splicing tend to be frameshifting both if they are not conserved at all (species-specific exons)²⁴⁰ and if they correspond to constitutively spliced exons (conserved exons with nonconserved alternative splicing).²⁵¹

3.5. Evolution of Alternative Splicing in Insects and Plants

Much less is known about evolution of alternative splicing in other taxa. A study of alternative splicing in two fruit flies, *D. melanogaster* and *D. pseudoobscura*, and the malarial mosquito *Anopheles gambiae* demonstrated a similar trend: much less conservation of alternative exons and sites compared with constitutive ones.²⁶³ Among *D. melanogaster* protein-coding segments bounded by splice sites (such as exons or, for example, segments between two alternative donor splice sites), 97% of constitutive ones and 75–80% of alternative ones could be aligned to orthologous *D. pseudoobscura* genes; for *A. gambiae*, these values were about 75% and 45%, respectively. Notwithstanding technical problems with spliced alignment at relatively large evolutionary distances, especially at gene termini, and dynamic intron insertion/loss processes, these results clearly demonstrate higher turnover of alternative regions compared with constitutive ones. When various types of alternative splicing were considered separately, it turned out that mutually exclusive exons are as conserved as constitutive ones and tolerate almost no intron insertions: a natural explanation for that could be that such insertions would disrupt the regulation of exon selection. Alternative donor splice sites are less conserved than alternative acceptor splice sites. Finally, cassette exons were the least conserved in the *D. melanogaster*–*A. gambiae* comparison, whereas for the two fruit flies, the least conserved type of alternatives were retained introns. Clearly, these observations do not provide all the answers, and the problem needs to be revisited once multiple *Drosophila* genomes become available.

A comparison of two plant genomes, *Arabidopsis* and rice, demonstrated all the familiar patterns, despite the fact that the main type of alternative splicing in plants is intron retention.³⁹ About 40% of alternative splicing events produce potential targets for NMD. About 40% of alternatively spliced genes from one species were also alternatively spliced in the other one (based on EST data), although the fraction of exact conservation (the same type of alternative splicing affecting the same site or intron) was much lower. Among conserved retained introns, more than half produced candidate NMD-inducing isoforms upon inclusion.

4. Origin of Alternative Regions

4.1. Sources of New Isoforms

So what are the sources for new alternatives? Theoretically, one may list several possibilities, dependent on the type of an alternative. Mutually exclusive exons may naturally arise from exon duplication;^{264,265} cassette exons^{254,266} and alternative sites^{267–269} stem from exonization events; the latter frequently follows insertion of repeats containing cryptic splice sites. One may also imagine migration of sites into exons (intronization of exon sequences) and events that

Table 1. Organization of the Groups of Mutually Exclusive Exons in the *Dscam* Genes of Insects^a

number of exons in other species	exon number (in <i>D. melanogaster</i>)				potential number of isoforms
	4	6	9	17	
<i>D. melanogaster</i>	12	48	33	2	38016
<i>D. pseudoobscura</i>	12	48	32	2	36864
<i>D. virilis</i>	12	52	32	1	19968
<i>A. gambiae</i> (malarial mosquito)	14	30	38	2	31920
<i>A. melifera</i> (honey bee)	8	45	17	2	12240

^a Data from ref 272.

would weaken or strengthen splice sites and thus influence retained introns, fixing one of two alternative variants: constitutively retained or spliced introns. The latter types have not been considered in detail yet. It should be noted, however, that subsequent intron insertion and loss may transform one type of alternative into another.²⁶³

4.2. Duplicated Exons

Many well-known cases of alternative splicing involve homologous, mutually exclusive exons; the most notorious one probably is the *Dscam* gene theoretically capable of forming (in *D. melanogaster*) 38 016 isoforms arising from four groups of mutually exclusive exons²⁷⁰ (exons 4, 6, 9, and 17, Table 1). This gene drives wiring of neuron connections in the neural system, and alternative splicing is an essential component of this process.²⁷¹ *Dscam* is conserved in other insects²⁷² retaining the overall organization, whereas in vertebrates, the corresponding exons are absent or present in single copies and thus constitutive.²⁷³ Construction of phylogenetic trees revealed both considerable conservation (exons 4 and 17, the latter encoding the transmembrane domain) and high diversity and turnover (exons 6 and 9).

Human, fruit fly *D. melanogaster*, and nematode *C. elegans* genomes were exhaustively analyzed in ref 265. About 11% of human genes and 7–8% of invertebrate genes were found to contain duplicated exons. The lengths of about 60% of duplicated exons were not divisible by 3, meaning that they are likely to be mutually exclusive: otherwise insertion of an upstream exon would shift the reading frame of a downstream exon. Indeed, analysis of EST data demonstrated that simultaneous incorporation of duplicated exons was extremely rare.

In a study of protein sequences from the Swiss-Prot database, alternatives leading to “substitution” variants (where two different amino acid fragments may occur in one position in alternative proteins) in about 10% of cases were caused by duplications;²⁶⁴ note that substitution variants are not necessarily mutually exclusive exons: a combination of an alternative splice site and a cassette exon, if they are consistently used in different isoforms, may generate exactly the same combination of observed proteins. Most exon duplications occurred before the radiation of mammalian orders and even vertebrates.

In some cases, it seems that homologous proteins utilize this diversity-generating mechanism independently in different evolutionary lineages. One example is the *Dscam* gene mentioned above, although in that case it is likely that the first duplications giving rise to the extant groups of homologous, alternatively spliced exons occurred once in the common ancestor of studied insects (although not necessarily

the last common ancestor: genomes of additional insects and arthropods are needed to elucidate this). Another example is more interesting: in three ion channel families, exon duplications occurred independently in the nematode, insect, and mammalian lineages.²⁷⁴ Notably, in many such cases the use of alternative exons leads to measurable changes in the ion channel properties; many examples are listed in ref 121.

A specific case is duplication or, rather, multiplication of first exons and the adjacent promoter regions.⁵⁶ The pattern of duplications may be different in human and rodent genes (plectin, GR), or even between mouse and rat (UGT1), and may involve untranslated exons⁵⁶ (NOS1). The duplications may occur within a very short time period, as demonstrated by the analysis of the human cancer/testis antigen family MAGE-A that contains several genes duplicated independently and repeatedly in the human and mouse lineages.²⁷⁵ One human gene from this family, MAGE-A4, has about ten very similar untranslated first exons (some copies seem to be incomplete), whereas other genes, most of which also are alternatively spliced in the 5'-UTR region, have only one copy of this exon.²⁷⁶ At least some variants of the first exons are transcribed and seem to be tissue-specific.²⁷⁷

4.3. Exonization

Another way to generate new, alternatively spliced exons is exonization of noncoding regions. To distinguish exonization from loss of exons, one can use alignment to bacterial and yeast orthologs that would serve as outgroups. One such study demonstrated that in about one-third of alternative protein regions, the ancestral state is the shorter isoform.²⁶⁶

Other outgroups also can be used. Large-scale studies of exons specific to the rodent lineage (compared with human) demonstrated that most of them are alternatively spliced (and rare, see above). In one study, rodent-specific exons rarely had homologous regions elsewhere in the genome; in particular, only a few originated from transposable elements,²⁵⁴ although another study showed that SINE retrotransposons, namely, B4 and MIR elements, are often involved in exonization-type alternative splicing.²⁷⁸ As usual, one possible cause for this discrepancy could be the (computationally defined) functionality, or lack thereof, of repeat-generated isoforms.

The already mentioned MAGE-A family of cancer/testis antigens provided an example where the events leading to emergence of new exons, all of which are alternatively spliced, could be studied in detail.²⁷⁶ The ancestor of this family was an intronless gene introduced by retroposition of a spliced mRNA, and since this had happened relatively recently, the sequence of the ancestral gene could be reconstituted. When this sequence was compared with the extant genes, it turned out that in many cases new exons had splice sites arising from point mutations at the one end and pre-existing cryptic sites at the other end. This observation agrees with the fact that in many cases of genetic disease caused by mutation in a splice site, a cryptic site is activated nearby and shows that the genome has ample opportunity for new splicing events.^{279–283}

Another special case is the emergence of alternative NAGNAG acceptor splice sites.²⁸⁴ Human single nucleotide polymorphisms (SNPs) involving NAGNAG sites were compared with the chimpanzee genome, and about 70% of alleles that could lead to alternative splicing (both N positions are not G) were found to be gained in human; in 16% of cases, evidence of alternative splicing was found in EST

databases. The comparison of the SNP density in the NAGNAG acceptor sites with a random model of SNP distribution in acceptor sites demonstrated that a surprisingly large fraction of young NAGNAG-containing alleles is explained by purifying selection against the NAGNAG sites.

A well-studied case of exonization is introduction of new splice sites by *Alu* repeats in the human genome. Indeed, analysis of EST data demonstrates that about 4% of proteins may have repeat-derived fragments. When only known functional proteins were considered, this fraction decreased to as low as 0.1%, although this latter value may be an underestimate due to the ascertainment bias caused by selection of well-studied proteins for the analysis.²⁸⁵ All known *Alu*-derived sites are alternative, and the molecular events leading to the use of cryptic splice sites in *Alu* repeats have been studied in detail, both computationally and in experiment.^{267–269} *Alu* repeat-derived splice sites not only extend existing exons but create cassette exons as well, although their functionality is supported only by the human–chimpanzee conservation and thus is not very reliable.¹⁷⁷

Point mutations that create new splice sites usually involve creation of GT (for donor sites) and AG (for acceptor sites) dinucleotides, whereas the mutations that guide alternative site choice and, in experiment, change the spliceosome preferences or even create constitutive sites may involve other consensus positions. In the human prelamina A recognition factor gene *NARF*, alternative splicing of an *Alu*-derived exon is provided by a tissue-dependent adenosine-to-inosine editing event that creates a splice site.⁶³ At that, it should be noted that human transcripts are edited more often than rodent, chicken, and fly transcripts²⁸⁶ and that the majority of edited sites have been observed within *Alu* sequences.^{287,288} Finally, mutations that create strong sites may lead to constitutive splicing and be highly deleterious.^{289–291}

Sequencing of several genes with *Alu*-derived exonized sequences, both exon extensions and cassette exons, in multiple primates showed how these events occurred in evolution.^{292–294} Exonization caused by creation of splice sites may occur millions of years after *Alu* insertion and may be reversed. Exonized fragments of *Alu* may disrupt the reading frame, and subsequent insertions or deletions may restore it. Although most such isoforms seem to be expressed at a low level and thus not affect the production of the protein, they still may confer selective disadvantage if the alternative isoform interferes with the protein functions.²⁹⁵ On the other hand, the fact that the exonized *Alu* fragments persist in different lineages and some of them contain only in-frame deletions shows that they may have a functional role.

Sometimes alternative splicing produces downstream regions that are read in two or even three reading frames. About 7% of human alternatively spliced genes contain such regions that are normally rather short²⁹⁶ (average length 133 nucleotides, close to the average exon length). Comparison with genomic sequences of other vertebrates demonstrates that such events are usually rather recent (primate-specific), as in orthologous regions of other species one of the reading frames is interrupted by a stop codon (about 50% of regions in rodents and about 80% in chicken). Although at the first glance such result is exactly what could be expected from simply aberrant splicing causing downstream frameshifts, this does not seem to be the case, because the study was conducted on a set of well-validated alternatively spliced genes. The functional advantages of such arrangement are

not clear, and one possibility is that they simply represent evolutionary deadlocks. If so, such a deadlock could be resolved by duplication of the exon read in several reading frames, leading to creation of mutually exclusive, independently evolving exons. It might be interesting to search for such examples, although clear-cut cases probably would be rare.

5. Changes in Alternative Splicing after Gene Duplication

Alternative splicing is just one mechanism of generating protein diversity, the other one being duplication of genes followed by substitutions, insertions, deletions, etc.²⁹⁷ By analogy to the situation with tissue-specific transcription and alternative splicing that have been shown to be independent,²⁵¹ one could expect independence between gene duplication and alternative splicing. But this does not seem to be the case.

In a study of human gene families duplicated at several time periods defined by splits of the human lineage with mouse, pufferfish *Fugu rubripes*, fruit fly *D. melanogaster*, and yeast *S. cerevisiae*, the fraction of alternatively spliced genes monotonically decreased in larger families.²⁹⁸ A possible explanation that some types of genes tend to diversify by duplication while others prefer alternative splicing has been rejected, because the prevalence of alternative splicing of singletons whose orthologs are duplicated in other species is similar to that of singletons corresponding to singletons. On the other hand, the situation is not absolutely clear-cut, since duplicated genes with singleton orthologs demonstrate higher prevalence of alternative splicing than duplicated genes whose orthologs are also duplicated. Finally, it has been mentioned that the observed difference should not be due to isoform loss in duplicated genes, since the fraction of alternatively spliced genes was smallest among recently duplicated genes and largest among ancient families.

In another study, not only the fact of alternative splicing but also the number of isoforms was taken into account; the latter was computed by aligning ESTs and mRNAs to the genomic sequence and counting the number of alignments producing a list of exons different from the longest RefSeq isoform.²⁹⁹ Given that most ESTs do not span an entire gene but still may cover several adjacent alternative regions, the obtained estimate is something between the number of isoforms and the number of elementary alternatives. Gene families were identified using simple BLAST-based clustering. Duplicated genes were shown to have less alternative isoforms than singletons; both the average number of isoforms and the fraction of alternatively spliced genes were relatively small for large gene families (>10 members). Using the protein identity as a proxy for the duplication time, the authors concluded that the fraction of alternatively spliced genes is lower among recently duplicated genes compared with more ancient duplications. Similar results were obtained on mouse, *D. melanogaster*, and *C. elegans* genes. Finally, the analysis of human genes duplicated after the human–mouse split demonstrated that these genes are rarely alternatively spliced compared with singletons.

Thus these two studies produced similar results. However, the interpretation of them was different. One possibility is the preferred loss of alternative isoforms immediately after duplication, which is a form of subfunctionalization, that is, differential loss of functions in duplicated copies.²⁹⁹ On the

other hand, this scenario does not explain why there are more alternatively spliced genes among ancient duplicates, unless one invokes gain of alternative splicing after the initial postduplication loss. The alternative is simply that alternative splicing and duplication are independent, balanced mechanisms for generating protein diversity, so that an alternative isoform is functionally equivalent to a paralog.²⁹⁸ To determine what model is correct, one would need to reconstruct the (alternative) exon–intron structures of the ancestral preduplication genes or, at least, directly compare the structures of duplicated genes to each other and to an outgroup genome (say, mouse or dog for human genes). Another hypothesis that needs to be checked is whether there is a correlation between alternative splicing in one species and duplication in another species. In particular, an almost direct test of the splicing-related subfunctionalization could be provided by comparing mammalian genes with genes of teleost fishes that have experienced whole-genome duplication followed by rapid purging of most duplicated copies.^{300–303} If the subfunctionalization explanation is correct, one could expect that fish orthologs of alternatively spliced human genes would more often retain both duplicated copies compared with orthologs of constitutively spliced genes.

In one more study, functional consequences of alternative splicing and gene duplication were considered.³⁰⁴ The anticorrelation between the fraction of alternatively spliced genes and the family size was observed both within the human genome and in the human–mouse and human–*Drosophila* comparisons. The degree of coexpression of alternatively spliced isoforms was similar to that of duplicated isoforms, and there was no anticorrelation in the expression patterns of alternatively spliced and duplicated isoforms. The functional consequences for the protein, naturally, were different: alternative splicing retains absolute identity in some (constitutive) regions, whereas the similarity between alternative regions is rather low, and the pattern of substitutions in duplicated genes is more uniform. The same holds for the insertion/deletion changes: the ones introduced by the alternative splicing are much larger than the indels in duplicated genes. There was little overlap between the indel positions in homologous alternatively spliced and duplicated isoforms; however, several interesting cases of parallelism were observed. While none of these findings are particularly surprising, they demonstrate that alternative splicing and gene duplication are independent mechanisms of generating sequence diversity.

6. Specifics of Selection in Alternatively Spliced Regions and Regulation of Alternative Splicing

6.1. Evolution Rates in Constitutive and Alternative Regions

Since the previously discussed studies seem to indicate that alternative regions often are relatively young, the next obvious question is what evolutionary forces shape them. Intuitively, one could expect that they would evolve faster than constitutive ones, both by an analogy with duplicated genes^{305–308} and by a simple assumption that a newly created isoform needs to adapt to its role. However, the analogy with duplications is not clear-cut, because recent studies seem to indicate that genes showing higher sequence conservation tend to be duplicated more often³⁰⁹ and that the overall rate of evolution of duplicated genes slows down despite the initial increase.³⁰⁵ The situation is further complicated by

regulatory sites, exonic splicing enhancers, and silencers that add an additional layer of selection.

In an early study of just 26 human/other mammal gene pairs, alternative regions were found to have a higher rate of nonsynonymous substitutions but a lower rate of synonymous ones, than constitutive regions.³¹⁰ The results of subsequent studies are summarized in Table 2. As above, most results were obtained for the human–mouse gene pairs, although other genome pairs (human–chimpanzee, mouse–rat, *Drosophila* spp.) were also considered.

The published values of the rate of nonsynonymous substitutions, K_a (aka K_n), and the rate of synonymous substitutions, K_s , are wildly different, both in constitutive exons and in alternative ones. There could be several reasons for that, described in subsequent paragraphs. Overall, it seems that K_a (alternative) tends to be larger than K_a (constitutive), although these observations are not very consistent. A slightly more consistent, although again nonuniversal observation, is that K_s (alternative) < K_s (constitutive).

Why are the published estimates of the evolutionary rates so different? First, the results may depend on whether the genomic conservation is considered sufficient or alternative exons are taken into account only if there is transcriptomic evidence for alternative splicing in both genomes.³¹¹ Second, the evolution rates are different for the major and minor isoform exons^{311,312} and for N-terminal, internal, and C-terminal alternative regions;^{313,314} the conservation is stronger for short exons and near splice sites.^{249,315,316} The latter fact demonstrates that splicing requirements significantly restrain the coding sequence and is consistent with the observation that the evolution rates increase in the regions of former exon junctions in intronless retroposed genes.³¹⁷ Third, the evolutionary rates in regions involved in simple alternative splicing events (cassette exons, alternative sites) and in complex ones are different: in a human–mouse comparison, simple regions had higher K_a , lower K_s , and, consequently, higher K_a/K_s than complex regions.³¹⁸ Similarly, the results on a control sample of constitutively spliced genes or exons may depend on its composition: as shown in a study that compared constitutive regions of alternatively spliced genes with conserved and nonconserved alternatives, the constitutive regions of genes with species-specific alternatives have higher synonymous and lower nonsynonymous evolution rates,³¹⁹ and in general, nonconserved exon–intron structure yields faster evolution of both constitutive and alternative exons.³²⁰ However, the latter observation is in conflict with the results of the analysis of several vertebrates: while both the synonymous substitution rate K_s and the K_a/K_s ratio are clearly higher in young exons, both constitutive and alternative, the remaining exons in genes with such exons show no difference in evolutionary rates.¹⁷⁷ Both nonsynonymous and especially synonymous substitution rates are lower in exons flanking long introns; this was ascribed to the presence of regulatory sites necessary to splice such introns.³²¹

Finally, even seemingly minor technical details might lead to different results. For instance, there are two ways of computing the averages: either to calculate the rates for the entire sample or to calculate the rates separately for each gene and then take the average. This may influence the results, because many individual exons are too short for meaningful calculations of evolutionary rates.^{313,314,322} Since the distributions of the evolutionary rates are nonsymmetric and have heavy tails, the median rates are lower than the averages,³¹² whereas usually just one of these values is

Table 2. Synonymous (K_s) and Nonsynonymous (K_a) Substitution Rates in Constant and Alternative Regions^a

ref	details	constant exons or regions			alternative exons or regions		
		K_a	K_s	$\omega = K_a/K_s$	K_a	K_s	$\omega = K_a/K_s$
311	cassette exons (major/medium/minor), genome-conserved	0.063	0.756	0.084	0.057/0.069/0.056	0.604/0.507/0.158	0.094/0.136/0.354
	cassette exons (major/medium/minor), transcriptome-conserved	0.053	0.748	0.071	0.053/0.055/0.059	0.402/0.189/0.117	0.131/0.291/0.500
	cassette exons (major/medium/minor), mouse–rat	0.018	0.189	0.097	0.017/0.020/0.024	0.147/0.114/0.046	0.117/0.180/0.523
	cassette exons (major/medium/minor), human–chimpanzee	0.0030	0.0145	0.206	0.0022/0.0036/0.0064	0.0122/0.0113/0.0089	0.182/0.318/0.716
315	cassette exons (human genome + rodent transcriptome)	0.103	0.311	0.331	0.058	0.115	0.504
313	gene regions (N-terminal/internal/C-terminal)	0.068	0.409	0.166	0.077 (0.075/0.074/0.133)	0.410 (0.404/0.421/0.446)	0.187 (0.186/0.176/0.297)
	gene regions (N-terminal/internal/C-terminal) slow-evolving genes	0.020	0.327	0.061	0.023	0.324	0.071
	gene regions (N-terminal/internal/C-terminal) medium-evolving genes	0.052	0.422	0.123	0.059	0.420	0.139
	gene regions (N-terminal/internal/C-terminal) fast-evolving genes	0.135	0.480	0.281	0.147	0.481	0.306
	cassette exons: major/minor or uncertain isoform	0.040	0.873	0.091	0.046/0.062	0.632/0.605	0.144/0.181
	cassette exons: major/minor medians	0.026	0.588	0.037	0.033/0.043	0.468/0.447	0.069/0.098
312	cassette exons: major/minor human–rat, medians	0.021	0.632	0.030	0.034/0.039	0.486/0.437	0.061/0.082
	cassette exons: major/minor mouse–rat, medians	0.007	0.192	0.016	0.014/0.019	0.181/0.179	0.070/0.092
249	cassette exons	0.081	0.510	0.158	0.065	0.166	0.390
314	<i>D. melanogaster</i> – <i>D. pseudoobscura</i> gene regions (N-terminal/internal/C-terminal)	0.22	0.27	0.81	0.25 (0.23/0.33/0.25)	0.31 (0.36/0.23/0.28)	0.80 (0.62/1.43/0.89)

^a Averages for human–mouse comparisons are given unless specified otherwise.

reported. The distribution of K_s (alternative) values is conspicuously bimodal,³²⁰ it might be a sign of two populations of alternative exons (one highly regulated and thus conserved; the other showing the same behavior as the constitutive exons) or simply due to annotation errors (for example, gene identification methods based on genomic comparison may mistakenly treat highly conserved intronic regions as candidate exons). Finally, all these analyses did not take into account the hypermutability of CpG dinucleotides, and this may confound the estimates of K_s and their interpretation.³²³

On the other hand, the results are largely consistent for the human–mouse (most studies), mouse–rat, human–chimpanzee, and fruit flies comparisons (Table 2). These observations mean that the alternative regions experience both positive selection on the protein level and stabilizing selection on the RNA level. The latter makes it difficult to apply the standard K_a/K_s test for positive selection, since it assumes that the synonymous sites evolve at a neutral rate. However, application of the McDonald–Kreitman test³²⁴ to human SNPs and human–chimpanzee substitutions demonstrated that about 9% of amino acid positions in minor isoform exons experience positive selection, and this fraction is as high as 25% for exons that are either conserved in the

mouse or dog genomes or belong to genes with high EST support.³²⁵ The major isoform exons and constitutive exons showed no signs of positive selection.

6.2. Regulatory Sites and Their Evolution

The difference in the synonymous evolution rate K_s between alternative and constitutive exons has been ascribed to the presence of regulatory sites in the former. Indeed, higher sequence conservation was observed near splice sites of alternative exons compared with constitutive ones in human–mouse,^{249,315,326} human–chimpanzee,³²⁷ and rice–wheat³²⁶ comparisons and in predicted exon splicing enhancers;³¹⁵ there are also fewer SNPs at exon termini.^{315,328} When regions of extreme conservation in synonymous codon positions between the human and mouse genomes were identified, more than 40% of them fell into alternative regions (compared with about 10% expected by chance), and these are enriched in candidate splicing enhancers and depleted for synonymous SNPs.³²⁹ Similarly, SNPs were found to be avoided in candidate exonic splicing enhancers,^{328,330} although it is not clear whether this affects predominantly synonymous codon positions³³⁰ or both synonymous and nonsynonymous positions.³²⁸

On the other hand, the above observations contradict the fact that the arrangement of candidate splicing enhancers is more conserved in constitutive exons compared with alternative exons.³²⁰ Further, when various features of alternative exons were analyzed separately, the most important factors influencing the K_a/K_s ratio were exon length and the fact that an exon overlaps a known protein domain, whereas the inclusion level and the exonic splicing enhancer density were less significant.³²² These observations are consistent with candidate exonic splicing enhancers³¹⁵ and silencers³³¹ being similarly abundant in constitutive and alternative exons, although the latter statement itself is subject to dispute, because more regulatory sites were observed in alternative exons in a number of studies.^{326,332,333} The resolution of the latter controversy may be that the relative abundance of splicing regulatory sites in alternative and constitutive exons may be different for different classes of regulatory elements.²⁷⁸

An observation that is not disputed is that both upstream and downstream intron regions adjacent to alternative exons are more conserved than introns adjacent to constitutive exons.^{241,249,334,335} The same holds for alternative donor and acceptor splice sites, although not for constitutive sites of the same exons.²⁴¹ The conservation is even higher for exons with strictly tissue-specific expression, and such exons are enriched in known and new regulatory motifs that dictate the tissue-dependent inclusion pattern.³³⁶ On the other hand, this effect is not limited to alternative splicing. Higher conservation is observed in introns of all tissue-specific genes compared with housekeeping ones,³³⁷ and intron regions adjacent to constitutive exons also tend to be conserved and enriched in candidate exonic splicing silencers, if the intron contains a strong cryptic splice site that could compete with the main one; this holds both for donor and acceptor sites.³³⁸

The average length of conserved intronic regions seems to be about 100–150 nucleotides,^{335,336} and this is consistent with the length of the region where intronic splicing enhancer sites seem to occur.³³⁹ In rodents, the evolutionary rate in regions immediately adjacent to splice sites and in first introns was found to be lower than that in 4-fold synonymous sites.³⁴⁰ A specific case is that of the NAGNAG alternative acceptor sites: the upstream intronic 30-nucleotide region is more conserved than the corresponding region of constitutive sites, although less conserved than the region upstream of cassette exons.³⁴¹ Similarly, higher conservation in adjacent exonic and intronic regions distinguishes alternative GYNGYN donor sites from constitutive ones.³⁴²

Unlike exonic enhancer motifs that are conserved in all vertebrate species, intronic splicing enhancers seem to be different in mammals and the *Fugu* fish.³⁴³ The candidate motifs were identified based on two premises: regulatory sites should be more frequent in the respective regions compared with intergenic controls, and they should be more abundant in the vicinity of weak splice sites.

On the other hand, some specific types of motifs, such as UGCAUG hexamers, which are binding sites for proteins of the Fox-1 family,³⁴⁴ were found to be conserved in intron regions adjacent to brain-specific exons in mammals, chicken, and fish.³⁴⁵ The latter study is a rare example of analysis of not only computationally predicted but also experimentally verified sites. Identification of pentamers and hexamers over-represented in conserved regions flanking cassette exons in two nematodes, *C. elegans* and *C. briggsae*, demonstrated

that (U)GCAUG may be a regulatory motif in these species as well, and the importance of this site for splicing of one gene, *unc-52* (containing 37 exons, three of which, exons 16, 17 and 18, are alternatively spliced) was confirmed in experiment.³⁴⁶ This is consistent with the fact that the FOX-1 protein of *C. elegans* is known to regulate splicing of the sex determination gene *xol-1*.³⁴⁷

Finally, in *Drosophila*, longer introns evolve more slowly than shorter introns³⁴⁸ and tend to flank alternatively spliced exons.³⁴⁹ One reason for that could be that long introns contain regulatory sites required either for efficient constitutive splicing of exons flanking such introns or for regulation of alternative splicing. An additional possibility is that some intronic conserved regions may serve as sites for recursive splicing.³⁵⁰ Candidate splice sites are over-represented in long introns and under-represented in short introns compared with control (antisense strand) and are conserved between two genomes of fruit flies, *D. melanogaster* and *D. pseudoobscura*. Four cases of iterative splicing were confirmed in experiment.

7. Recognition of Alternative Exons and Databases of Alternative Splicing

Many of the features of alternative exons and their evolution discussed above have been used to identify alternative exons and exonic/intronic splicing enhancers/silencers. Some methods are listed in Tables 3 and 4. We do not include methods based on EST analysis or spliced alignment of ESTs to the genome of origin. Most methods combine statistical features that distinguish alternative exons (specific oligonucleotide composition, weaker splice sites, length) with comparative data (relatively higher conservation, longer regions of conservation in adjacent introns).

8. Conclusion

Bioinformatics of splicing in general and comparative genomics of alternative splicing in particular look like pessimists' sciences. Neither donor and acceptor splice sites nor splicing enhancers and silencers can be reliably identified. No program can absolutely reliably predict even the exon–intron structure of a constitutively spliced gene, let alone all isoforms of an alternatively spliced gene, based on genomic sequence alone. This shows how little we know about sequence determinants of splicing and its regulation.

Not only fundamental problems, such as the introns-early/late theories or the functionality of NMD-inducing isoforms, but much more specific questions are subject to controversy. To name just a few: Is the prevalence of phase 0 introns due to older or younger introns? What are the rates of intron gain and loss in animals and in plants? What was the average number of introns in the last common ancestor of eukaryotes? What is the main mechanism of exon gain? What is the relationship between the exon inclusion level and frame preservation? Do minor form exons or NMD-inducing isoforms have a tendency for being tissue-specific? Is there evidence for subfunctionalization in alternative splicing of duplicated paralogs? Is there positive selection in alternative gene regions and stabilizing selection in synonymous codon positions of alternative exons, and if so, is the latter related to splicing enhancers? Are splicing enhancers more abundant in alternative exons compared with constitutive ones? What is the evolutionary range of exonic and intronic splicing

Table 3. Methods for Identification of Alternative Exons and Sites

ref	name	object	genome(s)	parameters/features	algorithmic technique
351	none	alternative vs constitutive exons	human and mouse	exon length, its divisibility by 3, sequence conservation in exon and adjacent introns, statistics of splice sites, triplet statistics in exon and intron	logical rules
352	none	alternative vs constitutive exons	human and mouse	exon length, its divisibility by 3, sequence conservation in exon and adjacent introns, statistics of splice sites, triplet statistics in exon and intron	support vector machine
252	ACESCAN	alternative vs constitutive exons	mammals (human and mouse)	exon and intron lengths, sequence conservation, oligonucleotide composition	regularized least-squares classifier
353	UNCOVER	cassette exons, exons with 5'- and 3'-extensions (alternative sites), retained introns vs conserved non-coding regions	human and mouse	substitution and deletion frequencies, splice site statistics, oligonucleotide composition	pair hidden Markov model
249	LL(E)	alternative vs constitutive exons	mammals	exon length, its divisibility by 3, splice site scores, sequence conservation in exon and adjacent introns	log-likelihood discriminator
253, 354	PSEP, ENACE	exons vs non-coding regions	mammals	sequence conservation in protein-coding regions	cross-species EST-genome and genome-genome spliced alignment with post-processing filters
355	Discipulus	alternative vs constitutive exons; retained vs constitutive introns	mammals	exon length, nucleotide composition	genetic programming
356	none	exons vs non-coding regions; cassette vs constitutive exons; retained introns	all genomes (human and mouse)	PFAM domains	spliced alignment of protein profiles to genome using the Viterbi algorithm (dynamic programming)
334	none	alternative vs constitutive exons	<i>D. melanogaster</i> , <i>D. pseudoobscura</i>	sequence conservation in exons and adjacent introns	logical rules applied to genomic alignment by VISTA
357	RASE	alternative exons vs introns and constant exons	<i>C. elegans</i>	oligonucleotide compositions, exon and intron lengths	support vector machine

enhancers? Does alternative splicing tend to avoid, shuffle, or disrupt protein domains? etc., etc.

Still, there is ground for cautious optimism as well, because despite all particular controversies, the emerging picture seems rather consistent. Splicing is noisy, both in a cell, due to spontaneous insertion or skipping of exons or using cryptic sites, and in evolution, since point mutations easily lead to new alternative splice sites or cassette exons. The damage from noise is controlled by the nonsense-mediated decay surveillance mechanism. On the other hand, this noise provides material for selection of new, beneficial variants.

A radical theory is that most new exons pass the stage of being alternative. That is, exonization due to emergence of a new splice site or duplication creates a new exon that initially exists as a minor isoform variant. If this isoform is NMD-inducing or functionally damaging, it is selected against; on the other hand, if this isoform has useful properties, it is fixed; its prevalence increases due to

additional mutations strengthening the sites, and, eventually, it becomes a major isoform or even a constitutive variant. At the same time, the new alternative region experiences both relaxed negative selection (since a minor isoform is not under a strong pressure) and positive selection (which fine-tunes its function). There also may be a flow of events in the opposite direction, with constitutive exons becoming cassette ones due to weakening of their sites. Minor isoform variants are not only easily born but also easily lost.

Of course, the reality is much more complicated than this simple and rather dogmatic schema. In particular, an important process leading to exon shuffling is large-scale genomic recombination, which creates chimeric genes encoding proteins with new combinations of domains and thus new functional properties. However, this is a completely different story. It might be interesting to estimate the relative importance of recombination and alternative splicing-related mechanisms of generation of new proteins. A technical problem will be that it involves events occurring at different

Table 4. Methods for Identification of Splicing Regulatory Sites (Exonic and Intronic Splicing Enhancers)

ref	name	object	genome(s)	basic idea	basic technique
358, 359	RESCUE-ESE	exonic splicing enhancers	mammals (human, mouse), fishes (<i>Fugu</i> , zebrafish)	ESEs are enriched in constitutive exons, especially with weak splice sites, and avoided in introns near splice sites	statistical analysis of hexamer frequencies in various gene regions
343	RESCUE-ISE	intronic splicing enhancers	human, mouse, fugu	ISEs are enriched in introns relative to exons and in introns with weak splice sites relative to introns with strong splice sites	statistical analysis of hexamer frequencies in various gene regions and clustering of candidate motifs
332, 338	ExonScan	exonic splicing silencers	mammals (human, mouse, rat, dog)	oligonucleotide motifs overrepresented in experimentally identified decanucleotides are enriched in exons with strong splice sites, alternatively spliced exons, and pseudoexons	experiment: fluorescence-activated screen
360, 361	ESEfinder	exonic splicing enhancers (binding sites for four SR proteins)	human	profiles (positional weight matrices) derived from direct experimental data	experiment: SELEX
346	Wobble Aware Bulk Aligner	intronic splicing enhancers	<i>C. elegans</i> and <i>C. briggsae</i>	ISEs are overrepresented and conserved in introns flanking cassette exons compared with all introns	hidden Markov model
362	none	exonic splicing enhancers and silencers	human	ESEs and ESSs are, respectively, enriched and depleted in internal non-coding exons compared with pseudoexons and 5'-UTRs of intronless genes	statistical analysis of octamer frequencies; clustering of candidate motifs; experimental validation of several randomly selected motifs in artificial splicing system
363	none	exonic splicing silencers	human	ESSs are enriched in pseudoexons compared with real exons or flanking intronic sequences	statistical analysis of hexamer; clustering of candidate motifs; experimental validation of the candidates

time scales, (fixed) recombinations being relatively rare and thus more ancient than new exons. At such distances we may not be able to analyze the fine details of alternative-splicing evolution.

If the “evolution by alternative splicing” theory is correct, analysis of multiple genomes and transcriptomes would allow us to see various stages of this process fixed in different species. If an exon is constitutive in one species and alternative in another one, it may be absent in an outgroup genome, which will indicate that it is a recent invention. Conversely, if a cassette exon is not conserved, it still may correspond to a constitutive exon in an outgroup species, thus showing that we observe a stage of gradual loss. These events should be accompanied with observable changes in the splice site strengths.

Conservation of a considerable fraction of NMD-inducing exons and sites demonstrates that they are also important. However, they still may not have an independent biological function of their own but rather serve as regulatory sinks, whose sole function is to be degraded by NMD if the produced transcript is no longer needed. If so, their pattern of evolution should be different; in particular, one might expect that such isoforms will not be subject to selection on the protein level but rather be selected on the RNA level, at least in splicing regulatory sites, enhancers, and silencers. Superficially, this may manifest as higher K_a and lower K_s compared with constitutive regions, and to distinguish this from bona fide positive selection, one needs to apply dedicated tests such as the McDonald–Kreitman test,

analysis of conserved regions within constitutive, translated alternative, and NMD-inducing exons, and their relationship with candidate splicing regulatory sites.

One of the lessons of this review is that many observations of specific facts taken out of a more general context may have different explanations. Thus we expect that in the near future the research focus will move toward greater integration of all aspects of alternative splicing and its evolution and thus resolve some of the seeming conflicts between results obtained in more narrow studies. It is important to consider separately old and new exons, exons that are constitutive in some species and alternative in others, frame-preserving and NMD-inducing variants, and major and minor isoforms and to distinguish between exonized and duplicated exons. It is also clear that to get a comprehensive picture, one needs to consider not only cassette exons but other types of elementary alternatives (mutually exclusive exons, alternative sites, retained introns). The evolutionary processes may be different for exons flanked by short introns and long ones, because the mechanisms of splicing might be different (intron and exon definition, respectively). Finally, it seems that changes in one region of a gene (e.g., exon gain) may influence evolution of other regions.

Of course, such studies will be done on multiple genomes. This would allow one to deduce the direction of evolution (e.g., exon gain vs loss) using outgroups, to distinguish between conserved (hence presumably functional) and species-specific (very young, maybe noisy) events, and to estimate the rate of evolutionary events. Examples of such

events are (lineage-specific) gain and loss of cassette exons and sites, synonymous and nonsynonymous substitutions, and SNPs. They have to be set in correspondence with the protein structure (do new exons arise between domains? are entire domains duplicated?) and for capability for protein coding or NMD induction, that is, frameshifting and in-frame stop codons. Further, these studies should benefit from availability of transcriptomic data in the form of ESTs, exonic, exon-junction, or tiling oligonucleotide arrays, that would allow one to identify exons and sites alternative in one species and constitutive in the other.

We believe that the next few years will provide some answers and, likely, set many more problems, so, studying of alternative splicing will continue to generate conflicting theories, bitter controversies, and new discoveries.

9. Acknowledgments

We are grateful to Ramil Nurtdinov and Vasily Ramensky for permission to cite their unpublished results. M.G. was partially supported by grants from the Howard Hughes Medical Institute (55005610), INTAS (05-1000008-8028), and the Russian Fund of Basic Research (07-04-00343). I.A. was supported by the *BioSapiens* project funded by the European Commission FP6 Programme, under the thematic area "Life sciences, genomics and biotechnology for health", contract number LHSG-CT-2003-503265. Both authors were partially supported by the Russian Academy of Sciences (program "Molecular and Cellular Biology").

10. References

- Berget, S. M.; Moore, C.; Sharp, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 3171–3175.
- Chow, L. T.; Gelinis, R. E.; Broker, T. R.; Roberts, R. J. *Cell* **1977**, *12*, 1–8.
- DeNoto, F. M.; Moore, D. D.; Goodman, H. M. *Nucleic Acids Res.* **1981**, *9*, 3719–3730.
- Adams, M. D.; Kerlavage, A. R.; Fields, C.; Venter, J. C. *Nat. Genet.* **1993**, *4*, 256–267.
- Asamizu, E.; Nakamura, Y.; Sato, S.; Tabata, S. *DNA Res.* **2000**, *7*, 175–180.
- Brentani, H.; Caballero, O. L.; Camargo, A. A.; da Silva, A. M.; da Silva, W. A., Jr.; Neto, E. D.; Grivet, M.; Gruber, A.; Guimaraes, P. E. M.; Hide, W.; Iseli, C.; Jongeneel, C. V.; Kelso, J.; Nagai, M. A.; Ojopi, E. P. B.; Osorio, E. C.; Reis, E. M. R.; Riggins, G. J.; Simpson, A. J. G.; de Souza, S.; Stevenson, B. J.; Strausberg, R. L.; Tajara, E. H.; Verjovski-Almeida, S.; Acencio, M. L.; Bengtson, M. H.; Bettoni, F.; Bodmer, W. F.; Briones, M. R. S.; Camargo, L. P.; Cavenee, W.; Cerutti, J. M.; Andrade, L. E. C.; dos Santos, P. C. C.; Costa, M. C. R.; da Silva, I. T.; Estecio, M. R. H.; Sa Ferreira, K.; Furnari, F. B.; Faria, M., Jr.; Galante, P. A. F.; Guimaraes, G. S.; Holanda, A. J.; Kimura, E. T.; Leerkes, M. R.; Lu, X.; Maciel, R. M. B.; Martins, E. A. L.; Massier, K. B.; Melo, A. S. A.; Mestriner, C. A.; Miracca, E. C.; Miranda, L. L.; Nobrega, F. G.; Oliveira, P. S.; Paquola, A. C. M.; Pandolfi, J. R. C.; Pardini, M. I. C.; Passeti, F.; Quackenbush, J.; Schnabel, B.; Sogayar, M. C.; Souza, J. E.; Valentini, S. R.; Zaiats, A. C.; Amaral, E. J.; Arnaldi, L. A. T.; de Araujo, A. G.; de Bessa, S. A.; Bicknell, D. C.; de Camaro, M. E. R.; Carraro, D. M.; Carrer, H.; Carvalho, A. F.; Colin, C.; Costa, F.; Curcio, C.; da Silva, I. D. C. G.; da Silva, N. P.; Dellamano, M.; El-Dorry, H.; Espreafico, E. M.; Ferreira, A. J. S.; Ferreira, C. A.; Fortes, M. A. H. Z.; Gama, A. H.; Giannella-Neto, D.; Giannella, M. L. C.; Giorgi, R. R.; Goldman, G. H.; Goldman, M. H. S.; Hackel, C.; Ho, P. L.; Kimura, E. M.; Kowalski, L. P.; Krieger, J. E.; Leite, L. C. C.; Lopes, A.; Luna, A. M. S. C.; Mackay, A.; Mari, S. K. N.; Marques, A. A.; Martins, W. K.; Montagnini, A.; Neto, M. M.; Nascimento, A. L. T. O.; Neville, A. M.; Nobrega, M. P.; O'Hare, M. J.; Otsuka, A. Y.; de Melo, A. I. R.; Paco-Larson, M. L.; Pereira, G. G.; da Silva, N. P.; Pesquero, J. B.; Pessoa, J. G.; Rahal, P.; Rainho, C. A.; Rodrigues, V.; Rogatto, S. R.; Romano, C. M.; Romeiro, J. G.; Rossi, B. M.; Rusticci, M.; de Sá, R. G.; Sant'Anna, S. C.; Sarmazo, M. L.; Silva, T. C. L.; Soares, F. A.; Sonati, M. F.; Souza, J. F.; Queiroz, D.; Valente, V.; Vettore, A. L.; Villanova, F. E.; Zago, M. A.; Zalberg, H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13418–13423.
- Clark, M. S.; Edwards, Y. J.; Peterson, D.; Clifton, S. W.; Thompson, A. J.; Sasaki, M.; Suzuki, Y.; Kikuchi, K.; Watabe, S.; Kawakami, K.; Sugano, S.; Elgar, G.; Johnson, S. L. *Genome Res.* **2003**, *13*, 2747–2753.
- Flinn, B.; Rothwell, C.; Griffiths, R.; Lague, M.; DeKoeper, D.; Sardana, R.; Audy, P.; Goyer, C.; Li, X. Q.; Wang-Pruski, G.; Regan, S. *Plant Mol. Biol.* **2005**, *59*, 407–433.
- Forment, J.; Gadea, J.; Huerta, L.; Abizanda, L.; Agusti, J.; Alamar, S.; Alos, E.; Andres, F.; Arribas, R.; Beltran, J. P.; Berbel, A.; Blazquez, M. A.; Brumos, J.; Canas, L. A.; Cercos, M.; Colmenero-Flores, J. M.; Conesa, A.; Estables, B.; Gandia, M.; Garcia-Martinez, J. L.; Gimeno, J.; Gisbert, A.; Gomez, G.; Gonzalez-Candelas, L.; Granell, A.; Guerri, J.; Lafuente, M. T.; Madueno, F.; Marcos, J. F.; Marques, M. C.; Martinez, F.; Martinez-Godoy, M. A.; Miralles, S.; Moreno, P.; Navarro, L.; Pallas, V.; Perez-Amador, M. A.; Perez-Valle, J.; Pons, C.; Rodrigo, I.; Rodriguez, P. L.; Royo, C.; Serrano, R.; Soler, G.; Tadeo, F.; Talon, M.; Terol, J.; Trenor, M.; Vaello, L.; Vicente, O.; Vidal, Ch.; Zacarias, L.; Conejero, V. *Plant Mol. Biol.* **2005**, *57*, 375–391.
- Kikuchi, S.; Satoh, K.; Nagata, T.; Kawagashira, N.; Doi, K.; Kishimoto, N.; Yazaki, J.; Ishikawa, M.; Yamada, H.; Ooka, H.; Hotta, I.; Kojima, K.; Namiki, T.; Ohneda, E.; Yahagi, W.; Suzuki, K.; Li, C. J.; Ohtsuki, K.; Shishiki, T.; Otomo, Y.; Murakami, K.; Iida, Y.; Sugano, S.; Fujimura, T.; Suzuki, Y.; Tsunoda, Y.; Kurosaki, T.; Kodama, T.; Masuda, H.; Kobayashi, M.; Xie, Q.; Lu, M.; Nariakawa, R.; Sugiyama, A.; Mizuno, K.; Yokomizo, S.; Niikura, J.; Ikeda, R.; Ishibiki, J.; Kawamata, M.; Yoshimura, A.; Miura, J.; Kusumegi, T.; Oka, M.; Ryu, R.; Ueda, M.; Matsubara, K.; Kawai, J.; Carninci, P.; Adachi, J.; Aizawa, K.; Arakawa, T.; Fukuda, S.; Hara, A.; Hashizume, W.; Hayatsu, N.; Imotani, K.; Ishii, Y.; Itoh, M.; Kagawa, I.; Kondo, S.; Konno, H.; Miyazaki, A.; Osato, N.; Ota, Y.; Saito, R.; Sasaki, D.; Sato, K.; Shibata, K.; Shinagawa, A.; Shiraki, T.; Yoshino, M.; Hayashizaki, Y.; Yasunishi, A. *Science* **2003**, *301*, 376–379.
- Lazo, G. R.; Chao, S.; Hummel, D. D.; Edwards, H.; Crossman, C. C.; Lui, N.; Matthews, D. E.; Carollo, V. L.; Hane, D. L.; You, F. M.; Butler, G. E.; Miller, R. E.; Close, T. J.; Peng, J. H.; Lapitan, N. L. V.; Gustafson, J. P.; Qi, L. L.; Echaliere, B.; Gill, B. S.; Dilbirligi, M.; Randhawa, H. S.; Gill, K. S.; Greene, R. A.; Sorrells, M. E.; Akhunov, E. D.; Dvorak, J.; Linkiewicz, A. M.; Dubcovsky, J.; Hossain, K. G.; Kalavacharla, V.; Kianian, S. F.; Mahmoud, A. A.; Miftahudin, M.; X.-F.; Conley, E. J.; Anderson, J. A.; Pathan, M. S.; Nguyen, H. T.; McGuire, P. E.; Qualset, C. O.; Anderson, O. D. *Genetics* **2004**, *168*, 585–593.
- Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M. C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; Kodzius, R.; Shimokawa, K.; Bajic, V. B.; Brenner, S. E.; Batalov, S.; Forrest, A. R. R.; Zavolan, M.; Davis, M. J.; Wilming, L. G.; Aidinis, V.; Allen, J. E.; Ambesi-Impombato, A.; Apweiler, R.; Aturaliya, R. N.; Bailey, T. L.; Bansal, M.; Baxter, L.; Beisel, K. W.; Bersano, T.; Bono, H.; Chalk, A. M.; Chiu, K. P.; Choudhary, V.; Christoffels, A.; Clutterbuck, D. R.; Crowe, M. L.; Dalla, E.; Dalrymple, B. P.; de Bono, B.; Della Gatta, G.; di Bernardo, D.; Down, T.; Engstrom, P.; Fagioli, M.; Faulkner, G.; Fletcher, C. F.; Fukushima, T.; Furuno, M.; Futaki, S.; Gariboldi, M.; Georgii-Hemming, P.; Gingeras, T. R.; Gojobori, T.; Green, R. E.; Gustincich, S.; Harbers, M.; Hayashi, Y.; Hensch, T. K.; Hirokawa, N.; Hill, D.; Huminecki, L.; Iacono, M.; Ikeo, K.; Iwama, A.; Ishikawa, T.; Jakt, M.; Kanapin, A.; Katoh, M.; Kawasawa, Y.; Kelso, J.; Kitamura, H.; Kitano, H.; Kollias, G.; Krishnan, S. P. T.; Kruger, A.; Kummerfeld, S. K.; Kurochkin, I. V.; Lareau, L. F.; Lazarevic, D.; Lipovich, L.; Liu, J.; Liuni, S.; McWilliam, S.; Madan Babu, M.; Madera, M.; Marchionni, L.; Matsuda, H.; Matsuzawa, S.; Miki, H.; Mignone, F.; Miyake, S.; Morris, K.; Mottagui-Tabar, S.; Mulder, N.; Nakano, N.; Nakauchi, H.; Ng, P.; Nilsson, R.; Nishiguchi, S.; Nishikawa, S.; Nori, F.; Ohara, O.; Okazaki, Y.; Orlando, V.; Pang, K. C.; Pavan, W. J.; Pavesi, G.; Pesole, G.; Petrosky, N.; Piazza, S.; Reed, J.; Reid, J. F.; Ring, B. Z.; Ringwald, M.; Rost, B.; Ruan, Y.; Salzberg, S. L.; Sandelin, A.; Schneider, C.; Schonbach, C.; Sekiguchi, K.; Sempile, C. A. M.; Seno, S.; Sessa, L.; Sheng, Y.; Shibata, Y.; Shimada, H.; Shimada, K.; Silva, D.; Sinclair, B.; Sperling, S.; Stupka, E.; Sugiura, K.; Sultana, R.; Takenaka, Y.; Taki, K.; Tammouja, K.; Tan, S. L.; Tang, S.; Taylor, M. S.; Tegner, J.; Teichmann, S. A.; Ueda, H. R.; van Nimwegen, E.; Verardo, R.; Wei, C. L.; Yagi, K.; Yamanishi, H.; Zabarovsky, E.; Zhu, S.; Zimmer, A.; Hide, W.; Bult, C.; Grimmond, S. M. T.; Teasdale, R. D.; Liu, E. T.; Brusic, V.; Quackenbush, J.; Wahlestedt, C.; Mattick, J. S.; Hume, D. A.; Kai, C.; Sasaki, D.; Tomaru, Y.; Fukuda, S.; Kanamori-Katayama, M.; Suzuki, M.; Aoki, J.; Arakawa, T.; Iida, J.; Imamura, K.; Itoh, M.; Kato, T.; Kawaji, H.; Kawagashira, N.; Kawashima, T.; Kojima, M.; Kondo, S.; Konno, H.; Nakano, K.; Ninomiya, N.; Nishio, T.; Okada, M.; Plessy, C.;

- Shibata, K.; Shiraki, T.; Suzuki, S.; Tagami, M.; Waki, K.; Watahiki, A.; Okamura-Oho, Y.; Suzuki, H.; Kawai, J.; Hayashizaki, Y. *Science* **2005**, *309*, 1559–1563.
- (13) Gerhard, D. S.; Wagner, L.; Feingold, E. A.; Shenmen, C. M.; Grouse, L. H.; Schuler, G.; Klein, S. L.; Old, S.; Rasooly, R.; Good, P.; Guyer, M.; Peck, A. M.; Derge, J. G.; Lipman, D.; Collins, F. S.; Jang, W.; Sherry, S.; Feolo, M.; Misquitta, L.; Lee, E.; Rotmistrovsky, K.; Greenhut, S. F.; Schaefer, C. F.; Buetow, K. H.; Bonner, T. I.; Haussler, D.; Kent, J.; Diekhans, M.; Furey, T.; Brent, M.; Prange, C.; Schreiber, K.; Shapiro, N.; Bhat, N. K.; Hopkins, R. F.; Hsie, F.; Driscoll, T.; Soares, M. B.; Bonaldo, M. F.; Casavant, T. L.; Scheetz, T. E.; Brownstein, M. J.; Usdin, T. B.; Toshiyuki, S.; Carninci, P.; Piao, Y.; Dudekula, D. B.; Ko, M. S. H.; Kawakami, K.; Suzuki, Y.; Sugano, S.; Gruber, C. E.; Smith, M. R.; Simmons, B.; Moore, T.; Waterman, R.; Johnson, S. L.; Ruan, Y.; Wei, C. L.; Mathavan, S.; Gunaratne, P. H.; Wu, J.; Garcia, A. M.; Hulyk, S. W.; Fuh, E.; Yuan, Y.; Sneed, A.; Kowis, C.; Hodgson, A.; Muzny, D. M.; McPherson, J.; Gibbs, R. A.; Fahey, J.; Helton, E.; Ketteman, M.; Madan, A.; Rodrigues, S.; Sanchez, A.; Whiting, M.; Madan, A.; Young, A. C.; Wetherby, K. D.; Granite, S. J.; Kwong, P. N.; Brinkley, C. P.; Pearson, R. L.; Bouffard, G. G.; Blakesly, R. W.; Green, E. D.; Dickson, M. C.; Rodriguez, A. C.; Grimwood, J.; Schmutz, J.; Myers, R. M.; Butterfield, Y. S. N.; Griffith, M.; Griffith, O. L.; Krzywinski, M. I.; Liao, N.; Morin, R.; Palmquist, D.; Petrescu, A. S.; Skalska, U.; Smailus, D. E.; Stott, J. M.; Schnerch, A.; Schein, J. E.; Jones, S. J. M.; Holt, R. A.; Baross, A.; Marra, M. A.; Clifton, S.; Makowski, K. A.; Bosak, S.; Malek, J. *Genome Res.* **2004**, *14*, 2121–2127.
- (14) Maeda, N.; Kasukawa, T.; Oyama, R.; Gough, J.; Frith, M.; Engstrom, P. G.; Lenhard, B.; Aturaliyar, R. N.; Batalov, S.; Beisel, K. W.; Bult, C. J.; Fletcher, C. F.; Forrest, A. R. R.; Furuno, M.; Hill, D.; Itoh, M.; Kanamori-Katayama, M.; Katayama, S.; Katoh, M.; Kawashima, T.; Quackenbush, J.; Ravasi, T.; Ring, B. Z.; Shibata, K.; Sugiura, K.; Takenaka, Y.; Teasdale, R. D.; Wells, C. A.; Zhu, Y.; Kai, C.; Kawai, J.; Hume, D. A.; Carninci, P.; Hayashizaki, Y. *PLoS Genet.* **2006**, *2*, e62.
- (15) Nunes, F. M.; Valente, V.; Sousa, J. F.; Cunha, M. A.; Pinheiro, D. G.; Maia, R. M.; Araujo, D. D.; Costa, M. C.; Martins, W. K.; Carvalho, A. F.; Monezi, N.; Nascimento, A. M.; Peixoto, P. M.; Silva, M. F.; Ramos, R. G.; Reis, L. F.; as-Neto, E.; Souza, S. J.; Simpson, A. J.; Zago, M. A.; Soares, A. E.; Bitondi, M. M.; Esprefaco, E. M.; Espindola, F. S.; Paco-Larson, M. L.; Simoes, Z. L.; Hartfelder, K.; Silva, W. A., Jr. *BMC Genomics* **2004**, *5*, 84.
- (16) Sogayar, M. C.; Camargo, A. A.; Bettioni, F.; Carraro, D. M.; Pires, L. C.; Parmigiani, R. B.; Ferreira, E. N.; de Sa, M. E.; do Rosario, D. d.; Simpson, A. J.; Cruz, L. O.; Degaki, T. L.; Festa, F.; Massirer, K. B.; Sogayar, M. C.; Filho, F. C.; Camargo, L. P.; Cunha, M. A.; de Souza, S. J.; Faria, M., Jr.; Giulianti, S.; Kopp, L.; de Oliveira, P. S.; Paiva, P. B.; Pereira, A. A.; Pinheiro, D. G.; Puga, R. D.; de Souza, J. E. S.; Albuquerque, D. M.; Andrade, L. E.; Baia, G. S.; Briones, M. R.; Cavaleiro-Luna, A. M.; Cerutti, J. M.; Costa, F. F.; Costanzi-Strauss, E.; Esprefaco, E. M.; Ferrasi, A. C.; Ferro, E. S.; Fortes, M. A.; Furchi, J. R.; Giannella-Neto, D.; Goldman, G. H.; Goldman, M. H.; Gruber, A.; Guimaraes, G. S.; Hackel, C.; Henrique-Silva, F.; Kimura, E. T.; Leoni, S. G.; Macedo, C.; Malnic, B.; Manzini, B. C.; Marie, S. K.; Martinez-Rossi, N. M.; Menossi, M.; Miracca, E. C.; Nagai, M. A.; Nobrega, F. G.; Nobrega, M. P.; Obashinjo, S. M.; Oliveira, M. K.; Orabona, G. M.; Otsuka, A. Y.; Paco-Larson, M. L.; Paixao, B. M.; Pandolfi, J. R.; Pardini, M. I.; Passos Bueno, M. R.; Passos, G. A.; Pesquero, J. B.; Pessoa, J. G.; Rahal, P.; Rainho, C. A.; Reis, C. P.; Ricca, T. I.; Rodrigues, V.; Rogatto, S. R.; Romano, C. M.; Romeiro, J. G.; Rossi, A.; Sa, R. G.; Sales, M. M.; Sant'Anna, S. C.; Santarosa, P. L.; Segato, F.; Silva, W. A., Jr.; Silva, I. D.; Silva, N. P.; Soares-Costa, A.; Sonati, M. F.; Strauss, B. E.; Tajara, E. H.; Valentini, S. R.; Villanova, F. E.; Ward, L. S.; Zanette, D. L. *Genome Res.* **2004**, *14*, 1413–1423.
- (17) Stapleton, M.; Carlson, J.; Brokstein, P.; Yu, C.; Champe, M.; George, R.; Guarin, H.; Kronmiller, B.; Pacleb, J.; Park, S.; Wan, K.; Rubin, G. M.; Celniker, S. E. *Genome Biol.* **2002**, *3*, research0080.
- (18) Sikder, S. K.; Kabat, E. A.; Morrison, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4045–4049.
- (19) Behlke, M. A.; Loh, D. Y. *Nature* **1986**, *322*, 379–382.
- (20) Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shohat, O.; Rotter, V. *Mol. Cell. Biol.* **1986**, *6*, 3232–3239.
- (21) Flanagan, J. G.; Chan, D. C.; Leder, P. *Cell* **1991**, *64*, 1025–1035.
- (22) Houck, K. A.; Ferrara, N.; Winer, J.; Cachianes, G.; Li, B.; Leung, D. W. *Mol. Endocrinol.* **1991**, *5*, 1806–1814.
- (23) Miki, T.; Bottaro, D. P.; Fleming, T. P.; Smith, C. L.; Burgess, W. H.; Chan, A. M.; Aaronson, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 246–250.
- (24) Naor, D.; Sionov, R. V.; Ish-Shalom, D. *Adv. Cancer Res.* **1997**, *71*, 241–319.
- (25) Schiaffino, S.; Reggiani, C. *Physiol. Rev.* **1996**, *76*, 371–423.
- (26) Adams, M. D.; Tarnig, R. S.; Rio, D. C. *Genes Dev.* **1997**, *11*, 129–138.
- (27) McKeown, M. *Curr. Opin. Genet. Dev.* **1992**, *2*, 299–303.
- (28) Saccone, G.; Pane, A.; Polito, L. C. *Genetica* **2002**, *116*, 15–23.
- (29) Sharp, P. A. *Cell* **1994**, *77*, 805–815.
- (30) Brett, D.; Hanke, J.; Lehmann, G.; Haase, S.; Delbruck, S.; Krueger, S.; Reich, J.; Bork, P. *FEBS Lett.* **2000**, *474*, 83–86.
- (31) Mironov, A. A.; Fickett, J. W.; Gelfand, M. S. *Genome Res.* **1999**, *9*, 1288–1293.
- (32) Brett, D.; Pospisil, H.; Valcarcel, J.; Reich, J.; Bork, P. *Nat. Genet.* **2002**, *30*, 29–30.
- (33) Kim, H.; Klein, R.; Majewski, J.; Ott, J. *Nat. Genet.* **2004**, *36*, 915–916.
- (34) Harrington, E. D.; Boue, S.; Valcarcel, J.; Reich, J. G.; Bork, P. *Nat. Genet.* **2004**, *36*, 916–917.
- (35) Kim, E.; Magen, A.; Ast, G. *Nucleic Acids Res.* **2007**, *35*, 125–131.
- (36) Kan, Z.; States, D.; Gish, W. *Genome Res.* **2002**, *12*, 1837–1845.
- (37) Iida, K.; Seki, M.; Sakurai, T.; Satou, M.; Akiyama, K.; Toyoda, T.; Konagaya, A.; Shinozaki, K. *Nucleic Acids Res.* **2004**, *32*, 5096–5103.
- (38) Campbell, M. A.; Haas, B. J.; Hamilton, J. P.; Mount, S. M.; Buell, C. R. *BMC Genomics* **2006**, *7*, 327.
- (39) Wang, B. B.; Brendel, V. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7175–7180.
- (40) Ner-Gaon, H.; Fluhr, R. *DNA Res.* **2006**, *13*, 111–121.
- (41) Ner-Gaon, H.; Halachmi, R.; Savaldi-Goldstein, S.; Rubin, E.; Ophir, R.; Fluhr, R. *Plant J.* **2004**, *39*, 877–885.
- (42) Ner-Gaon, H.; Leviatan, N.; Rubin, E.; Fluhr, R. *Plant Physiol.* **2007**.
- (43) Davis, C. A.; Grate, L.; Spingola, M.; Ares, M., Jr. *Nucleic Acids Res.* **2000**, *28*, 1700–1706.
- (44) Graveley, B. R. *Trends Genet.* **2001**, *17*, 100–107.
- (45) Neugebäude, C.; Chalvet, F.; Wincker, P.; Gaillardin, C.; Casaregola, S. *Eukaryotic Cell* **2005**, *4*, 615–624.
- (46) Okazaki, K.; Niwa, O. *DNA Res.* **2000**, *7*, 27–30.
- (47) Muhia, D. K.; Swales, C. A.; Eckstein-Ludwig, U.; Saran, S.; Polley, S. D.; Kelly, J. M.; Schaap, P.; Krishna, S.; Baker, D. A. *J. Biol. Chem.* **2003**, *278*, 22014–22022.
- (48) Harrison, P. M.; Kumar, A.; Lang, N.; Snyder, M.; Gerstein, M. *Nucleic Acids Res.* **2002**, *30*, 1083–1090.
- (49) Modrek, B.; Lee, C. *Nat. Genet.* **2002**, *30*, 13–19.
- (50) Anderson, C. L.; Zundel, M. A.; Werner, R. *Genomics* **2005**, *85*, 238–244.
- (51) Bourdon, J. C.; Fernandes, K.; Murray-Zmijewski, F.; Liu, G.; Diot, A.; Xirodimas, D. P.; Saville, M. K.; Lane, D. P. *Genes Dev.* **2005**, *19*, 2122–2137.
- (52) de la Grange, P.; Dutertre, M.; Martin, N.; Auboeuf, D. *Nucleic Acids Res.* **2005**, *33*, 4276–4284.
- (53) Irvin-Wilson, C. V.; Chaudhuri, G. *Breast Cancer Res.* **2005**, *7*, R563–R569.
- (54) Kitagawa, N.; Washio, T.; Kosugi, S.; Yamashita, T.; Higashi, K.; Yanagawa, H.; Higo, K.; Satoh, K.; Ohtomo, Y.; Sunako, T.; Murakami, K.; Matsubara, K.; Kawai, J.; Carninci, P.; Hayashizaki, Y.; Kikuchi, S.; Tomita, M. *Bioinformatics* **2005**, *21*, 1758–1763.
- (55) Tan, J. S.; Mohandas, N.; Conboy, J. G. *Blood* **2006**, *107*, 2557–2561.
- (56) Zhang, T.; Haws, P.; Wu, Q. *Genome Res.* **2004**, *14*, 79–89.
- (57) Nagasaki, H.; Arita, M.; Nishizawa, T.; Suwa, M.; Gotoh, O. *Gene* **2005**, *364*, 53–62.
- (58) Le Texier, V.; Riethoven, J. J.; Kumanduri, V.; Gopalakrishnan, C.; Lopez, F.; Gautheret, D.; Thanaraj, T. A. *BMC Bioinf.* **2006**, *7*, 169.
- (59) Tian, B.; Hu, J.; Zhang, H.; Lutz, C. S. *Nucleic Acids Res.* **2005**, *33*, 201–212.
- (60) Yan, J.; Marr, T. G. *Genome Res.* **2005**, *15*, 369–375.
- (61) Misra, S.; Crosby, M. A.; Mungall, C. J.; Matthews, B. B.; Campbell, K. S.; Hradecky, P.; Huang, Y.; Kaminker, J. S.; Millburn, G. H.; Prochnik, S. E.; Smith, C. D.; Tupy, J. L.; Whitfield, E. J.; Bayraktaroglu, L.; Berman, B. P.; Bettencourt, B. R.; Celniker, S. E.; de Grey, A. D.; Drysdale, R. A.; Harris, N. L.; Richter, J.; Russo, S.; Schroeder, A. J.; Shu, S. Q.; Stapleton, M.; Yamada, C.; Ashburner, M.; Gelbart, W. M.; Rubin, G. M.; Lewis, S. E. *Genome Biol.* **2002**, *3*, research0083.
- (62) Bornes, S.; Boulard, M.; Hieblot, C.; Zanibellato, C.; Iacovoni, J. S.; Prats, H.; Touriol, C. *J. Biol. Chem.* **2004**, *279*, 18717–18726.
- (63) Lev-Maor, G.; Sorek, R.; Levanon, E. Y.; Paz, N.; Eisenberg, E.; Ast, G. *Genome Biol.* **2007**, *8*, R29.
- (64) Rueter, S. M.; Dawson, T. R.; Emeson, R. B. *Nature* **1999**, *399*, 75–80.
- (65) Cuccurese, M.; Russo, G.; Russo, A.; Pietropaolo, C. *Nucleic Acids Res.* **2005**, *33*, 5965–5977.
- (66) Ni, J. Z.; Grate, L.; Donohue, J. P.; Preston, C.; Nobida, N.; O'Brien, G.; Shiue, L.; Clark, T. A.; Blume, J. E.; Ares, M., Jr. *Genes Dev.* **2007**, *21*, 708–718.
- (67) Ohler, U. Personal communication, 2006.

- (68) Hughes, T. A. *Trends Genet.* **2006**, *22*, 119–122.
- (69) Fededa, J. P.; Petrillo, E.; Gelfand, M. S.; Neverov, A. D.; Kadener, S.; Noguez, G.; Pelisch, F.; Baralle, F. E.; Muro, A. F.; Kornblihtt, A. R. *Mol. Cell* **2005**, *19*, 393–404.
- (70) Hicks, M. J.; Yang, C. R.; Kotlajich, M. V.; Hertel, K. J. *PLoS Biol.* **2006**, *4*, e147.
- (71) Cramer, P.; Caceres, J. F.; Cazalla, D.; Kadener, S.; Muro, A. F.; Baralle, F. E.; Kornblihtt, A. R. *Mol. Cell* **1999**, *4*, 251–258.
- (72) Kornblihtt, A. R. *Curr. Opin. Cell Biol.* **2005**, *17*, 262–268.
- (73) Akiva, P.; Toporik, A.; Edelman, S.; Peretz, Y.; Diber, A.; Shemesh, R.; Novik, A.; Sorek, R. *Genome Res.* **2006**, *16*, 30–36.
- (74) Parra, G.; Reymond, A.; Dabbouseh, N.; Dermitzakis, E. T.; Castelo, R.; Thomson, T. M.; Antonarakis, S. E.; Guigo, R. *Genome Res.* **2006**, *16*, 37–44.
- (75) Pacheco, T. R.; Gomes, A. Q.; Barbosa-Morais, N. L.; Benes, V.; Ansong, W.; Wollerton, M.; Smith, C. W.; Valcarcel, J.; Carmo-Fonseca, M. *J. Biol. Chem.* **2004**, *279*, 27039–27049.
- (76) Popielarz, M.; Cavaloc, Y.; Mattei, M. G.; Gattoni, R.; Stevenin, J. *J. Biol. Chem.* **1995**, *270*, 17830–17835.
- (77) Sampson, N. D.; Hewitt, J. E. *Gene* **2003**, *305*, 91–100.
- (78) Sureau, A.; Gattoni, R.; Dooghe, Y.; Stevenin, J.; Soret, J. *EMBO J.* **2001**, *20*, 1785–1796.
- (79) Iida, K.; Go, M. *Mol. Biol. Evol.* **2006**, *23*, 1085–1094.
- (80) Kalyana, M.; Lopato, S.; Voronin, V.; Barta, A. *Nucleic Acids Res.* **2006**, *34*, 4395–4405.
- (81) Le, K.; Mitsouras, K.; Roy, M.; Wang, Q.; Xu, Q.; Nelson, S. F.; Lee, C. *Nucleic Acids Res.* **2004**, *32*, e180.
- (82) Modrek, B.; Resch, A.; Grasso, C.; Lee, C. *Nucleic Acids Res.* **2001**, *29*, 2850–2859.
- (83) Xu, Q.; Modrek, B.; Lee, C. *Nucleic Acids Res.* **2002**, *30*, 3754–3766.
- (84) Pritsker, M.; Doniger, T. T.; Kramer, L. C.; Westcot, S. E.; Lemischka, I. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14290–14295.
- (85) Clark, T. A.; Schweitzer, A. C.; Chen, T. X.; Staples, M. K.; Lu, G.; Wang, H.; Williams, A.; Blume, J. E. *Genome Biol.* **2007**, *8*, R64.
- (86) Ye, G.; Holste, D.; Kreiman, G.; Burge, C. B. *Genome Biol.* **2004**, *5*, R74.
- (87) Johnson, J. M.; Castle, J.; Garrett-Engele, P.; Kan, Z.; Loerch, P. M.; Armour, C. D.; Santos, R.; Schadt, E. E.; Stoughton, R.; Shoemaker, D. D. *Science* **2003**, *302*, 2141–2144.
- (88) Pan, Q.; Shai, O.; Misquitta, C.; Zhang, W.; Saltzman, A. L.; Mohammad, N.; Babak, T.; Siu, H.; Hughes, T. R.; Morris, Q. D.; Frey, B. J.; Blencowe, B. J. *Mol. Cell* **2004**, *16*, 929–941.
- (89) Brett, D.; Kemmer, W.; Koch, G.; Roefzaad, C.; Gross, S.; Schlag, P. M. *Oncogene* **2001**, *20*, 4581–4585.
- (90) Hui, L.; Zhang, X.; Wu, X.; Lin, Z.; Wang, Q.; Li, Y.; Hu, G. *Oncogene* **2004**, *23*, 3013–3023.
- (91) Xie, H.; Zhu, W. Y.; Wasserman, A.; Grebinskiy, V.; Olson, A.; Mintz, L. *Genomics* **2002**, *80*, 326–330.
- (92) Xu, Q.; Lee, C. *Nucleic Acids Res.* **2003**, *31*, 5635–5643.
- (93) Gardina, P. J.; Clark, T. A.; Shimada, B.; Staples, M. K.; Yang, Q.; Veitch, J.; Schweitzer, A.; Awad, T.; Sugnet, C.; Dee, S.; Davies, C.; Williams, A.; Turpaz, Y. *BMC Genomics* **2006**, *7*, 325.
- (94) Adams, M.; Jones, J. L.; Walker, R. A.; Pringle, J. H.; Bell, S. C. *Cancer Res.* **2002**, *62*, 3289–3297.
- (95) Cragg, M. S.; Chan, H. T.; Fox, M. D.; Tutt, A.; Smith, A.; Oscier, D. G.; Hamblin, T. J.; Glennie, M. J. *Blood* **2002**, *100*, 3068–3076.
- (96) Feltes, C. M.; Kudo, A.; Blaschuk, O.; Byers, S. W. *Cancer Res.* **2002**, *62*, 6688–6697.
- (97) Koslowski, M.; Tureci, O.; Bell, C.; Krause, P.; Lehr, H. A.; Brunner, J.; Seitz, G.; Nestle, F. O.; Huber, C.; Sahin, U. *Cancer Res.* **2002**, *62*, 6750–6755.
- (98) Saito, H.; Nakatsuru, S.; Inazawa, J.; Nishihira, T.; Park, J. G.; Nakamura, Y. *Oncogene* **1997**, *14*, 617–621.
- (99) Stimpfl, M.; Tong, D.; Fasching, B.; Schuster, E.; Obermair, A.; Leodolter, S.; Zeillinger, R. *Clin. Cancer Res.* **2002**, *8*, 2253–2259.
- (100) Kalnina, Z.; Zayakin, P.; Silina, K.; Line, A. *Genes, Chromosomes Cancer* **2005**, *42*, 342–357.
- (101) Srebrow, A.; Kornblihtt, A. R. *J. Cell Sci.* **2006**, *119*, 2635–2641.
- (102) Venables, J. P. *Bioessays* **2006**, *28*, 378–386.
- (103) Atanelov, L.; Xu, Q.; Rad, R.; Lee, C. *J. Gastroenterol.* **2005**, *40* (Suppl 16), 14–20.
- (104) Venables, J. P. *Cancer Res.* **2004**, *64*, 7647–7654.
- (105) Hagiwara, M. *Biochim. Biophys. Acta* **2005**, *1754*, 324–331.
- (106) Grahm, R. R.; Kozlyev, S. V.; Baechler, E. C.; Reddy, M. V. P.; Plenge, R. M.; Bauer, J. W.; Ortmann, W. A.; Koeuth, T.; Gonzalez Escribano, M. F. G.; Pons-Estel, B.; Petri, M.; Daly, M.; Gregersen, P. K.; Martin, J.; Altschuler, D.; Behrens, T. W.; Alarcon-Riquelme, M. E. *Nat. Genet.* **2006**, *38*, 550–555.
- (107) Hollanda, L. M.; Lima, C. S.; Cunha, A. F.; Albuquerque, D. M.; Vassallo, J.; Ozelo, M. C.; Joazeiro, P. P.; Saad, S. T.; Costa, F. F. *Nat. Genet.* **2006**, *38*, 807–812.
- (108) Ueda, H.; Howson, J. M.; Esposito, L.; Heward, J.; Snook, H.; Chamberlain, G.; Rainbow, D. B.; Hunter, K. M.; Smith, A. N.; Di Genova, G.; Herr, M. H.; Dahlman, I.; Payne, F.; Smyth, D.; Lowe, C.; Twells, R. C.; Howlett, S.; Healy, B.; Nutland, S.; Rance, H. E.; Everett, V.; Smink, L. J.; Lam, A. C.; Cordell, H. J.; Walker, N. M.; Bordin, C.; Hulme, J.; Motzo, C.; Cucca, F.; Hess, J. F.; Metzker, M. L.; Rogers, J.; Gregory, S.; Allahabadia, A.; Nithiyanathan, R.; Tuomilehto-Wolf, E.; Tuomilehto, J.; Bingley, P.; Gillespie, K. M.; Undlien, D. E.; Ronningen, K. S.; Guja, C.; Ionescu-Tirgoviste, C.; Savage, D. A.; Maxwell, A. P.; Carson, D. J.; Patterson, C. C.; Franklyn, J. A.; Clayton, D. G.; Peterson, L. B.; Wicker, L. S.; Todd, J. A.; Gough, S. C. L. *Nature* **2003**, *423*, 506–511.
- (109) Tate, S. K.; Depondt, C.; Sisodiya, S. M.; Cavalleri, G. L.; Schorge, S.; Soranzo, N.; Thom, M.; Sen, A.; Shorvon, S. D.; Sander, J. W.; Wood, N. W.; Goldstein, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5507–5512.
- (110) Homma, K.; Kikuno, R. F.; Nagase, T.; Ohara, O.; Nishikawa, K. *J. Mol. Biol.* **2004**, *343*, 1207–1220.
- (111) Kriventseva, E. V.; Koch, I.; Apweiler, R.; Vingron, M.; Bork, P.; Gelfand, M. S.; Sunyaev, S. *Trends Genet.* **2003**, *19*, 124–128.
- (112) Taneri, B.; Snyder, B.; Novoradovsky, A.; Gaasterland, T. *Genome Biol.* **2004**, *5*, R75.
- (113) Cline, M. S.; Shigeta, R.; Wheeler, R. L.; Siani-Rose, M. A.; Kulp, D.; Loraine, A. E. *Pac. Symp. Biocomput.* **2004**, 17–28.
- (114) Stetefeld, J.; Ruegg, M. A. *Trends Biochem. Sci.* **2005**, *30*, 515–521.
- (115) Davis, M. J.; Hanson, K. A.; Clark, F.; Fink, J. L.; Zhang, F.; Kasukawa, T.; Kai, C.; Kawai, J.; Carninci, P.; Hayashizaki, Y.; Teasdale, R. D. *PLoS Genet.* **2006**, *2*, e46.
- (116) Wang, P.; Yan, B.; Guo, J. T.; Hicks, C.; Xu, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18920–18925.
- (117) Romero, P. R.; Zaidi, S.; Fang, Y. Y.; Uversky, V. N.; Radivojac, P.; Oldfield, C. J.; Cortese, M. S.; Sickmeier, M.; LeGall, T.; Obradovic, Z.; Dunker, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8390–8395.
- (118) Yura, K.; Shionyu, M.; Hagino, K.; Hijikata, A.; Hirashima, Y.; Nakahara, T.; Eguchi, T.; Shinoda, K.; Yamaguchi, A.; Takahashi, K.; Itoh, T.; Imanishi, T.; Gojbori, T.; Go, M. *Gene* **2006**, *380*, 63–71.
- (119) Liu, S.; Altman, R. B. *Nucleic Acids Res.* **2003**, *31*, 4828–4835.
- (120) Resch, A.; Xing, Y.; Modrek, B.; Gorlick, M.; Riley, R.; Lee, C. J. *Proteome Res.* **2004**, *3*, 76–83.
- (121) Stamm, S.; Ben-Ari, S.; Rafalska, I.; Tang, Y.; Zhang, Z.; Toiber, D.; Thanaraj, T. A.; Soreq, H. *Gene* **2005**, *344*, 1–20.
- (122) Xing, Y.; Xu, Q.; Lee, C. *FEBS Lett.* **2003**, *555*, 572–578.
- (123) Loraine, A. E.; Helt, G. A.; Cline, M. S.; Siani-Rose, M. A. *J. Bioinform. Comput. Biol.* **2003**, *1*, 289–306.
- (124) Li, J.; Li, X.; Guo, L.; Lu, F.; Feng, X.; He, K.; Wei, L.; Chen, Z.; Qu, L. J.; Gu, H. *J. Exp. Bot.* **2006**, *57*, 1263–1273.
- (125) Hamilton, A. T.; Huntley, S.; Tran-Gyamfi, M.; Baggott, D. M.; Gordon, L.; Stubbs, L. *Genome Res.* **2006**, *16*, 584–594.
- (126) Huntley, S.; Baggott, D. M.; Hamilton, A. T.; Tran-Gyamfi, M.; Yang, S.; Kim, J.; Gordon, L.; Branscomb, E.; Stubbs, L. *Genome Res.* **2006**, *16*, 669–677.
- (127) Forrest, A. R.; Taylor, D. F.; Crowe, M. L.; Chalk, A. M.; Waddell, N. J.; Kollé, G.; Faulkner, G. J.; Kodzius, R.; Katayama, S.; Wells, C.; Kai, C.; Kawai, J.; Carninci, P.; Hayashizaki, Y.; Grimmond, S. M. *Genome Biol.* **2006**, *7*, R5.
- (128) Milanese, L.; Petrillo, M.; Sepe, L.; Boccia, A.; D'Agostino, N.; Passamano, M.; Di Nardo, S.; Tasco, G.; Casadio, R.; Paoletta, G. *BMC Bioinf.* **2005**, *6* (Suppl 4), S20.
- (129) Blencowe, B. J. *Cell* **2006**, *126*, 37–47.
- (130) Florea, L. *Briefings Bioinf.* **2006**, *7*, 55–69.
- (131) Gelfand, M. S. *Handbook of Computational Molecular Biology*; Chapman & Hall/CRC: Boca Raton, FL, 2006; Chapter 2, pp 2-1–2-18.
- (132) Lee, C.; Roy, M. *Genome Biol.* **2004**, *5*, 231.
- (133) Matlin, A. J.; Clark, F.; Smith, C. W. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 386–398.
- (134) Lee, C.; Wang, Q. *Briefings Bioinf.* **2005**, *6*, 23–33.
- (135) Zavolan, M.; van Nimwegen, E. *Curr. Opin. Struct. Biol.* **2006**, *16*, 362–367.
- (136) Bracco, L.; Throo, E.; Cochet, O.; Einstein, R.; Maurier, F. *Prog. Mol. Subcell. Biol.* **2006**, *44*, 1–25.
- (137) Cuperlovic-Culf, M.; Belacel, N.; Culf, A. S.; Ouellette, R. J. *Drug Discovery Today* **2006**, *11*, 983–990.
- (138) Srinivasan, K.; Shiue, L.; Hayes, J. D.; Centers, R.; Fitzwater, S.; Loewen, R.; Edmondson, L. R.; Bryant, J.; Smith, M.; Rommelfanger, C.; Welch, V.; Clark, T. A.; Sugnet, C. W.; Howe, K. J.; Mandel-Gutfreund, Y.; Ares, M. *Methods* **2005**, *37*, 345–359.
- (139) Ast, G. *Sci. Am.* **2005**, *292*, 40–47.
- (140) Downes, S. M. *Hist. Philos. Life Sci.* **2004**, *26*, 91–104.
- (141) Galperin, M. Y. *Nucleic Acids Res.* **2006**, *34*, D3–D5.

- (142) Ast, G. *Nat. Rev. Genet.* **2004**, *5*, 773–782.
- (143) Boue, S.; Letunic, I.; Bork, P. *Bioessays* **2003**, *25*, 1031–1034.
- (144) Lareau, L. F.; Green, R. E.; Bhatnagar, R. S.; Brenner, S. E. *Curr. Opin. Struct. Biol.* **2004**, *14*, 273–282.
- (145) King, Y.; Lee, C. *Nat. Rev. Genet.* **2006**, *7*, 499–509.
- (146) Koonin, E. V. *Biol. Direct* **2006**, *1*, 22.
- (147) Rogozin, I. B.; Sverdlov, A. V.; Babenko, V. N.; Koonin, E. V. *Briefings Bioinf.* **2005**, *6*, 118–134.
- (148) Roy, S. W.; Gilbert, W. *Nat. Rev. Genet.* **2006**, *7*, 211–221.
- (149) Blake, C. C. F. *Nature* **1978**, *273*, 267.
- (150) Doolittle, W. F. *Nature* **1978**, *272*, 581–582.
- (151) Gilbert, W. *Nature* **1978**, *271*, 501.
- (152) Cavalier-Smith, T. *Nature* **1985**, *315*, 283–284.
- (153) Cavalier-Smith, T. *Trends Genet.* **1991**, *7*, 145–148.
- (154) Palmer, J. D.; Logsdon, J. M., Jr. *Curr. Opin. Genet. Dev.* **1991**, *1*, 470–477.
- (155) Sharp, P. A. *Cell* **1985**, *42*, 397–400.
- (156) Stoltzfus, A.; Spencer, D. F.; Zuker, M.; Logsdon, J. M., Jr.; Doolittle, W. F. *Science* **1994**, *265*, 202–207.
- (157) Stoltzfus, A. *Nature* **1994**, *369*, 526–527.
- (158) Cho, G.; Doolittle, R. F. *J. Mol. Evol.* **1997**, *44*, 573–584.
- (159) Dibb, N. J.; Newman, A. J. *EMBO J.* **1989**, *8*, 2015–2021.
- (160) Fedorov, A.; Merican, A. F.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16128–16133.
- (161) Frugoli, J. A.; McPeck, M. A.; Thomas, T. L.; McClung, C. R. *Genetics* **1998**, *149*, 355–365.
- (162) Hartung, F.; Blattner, F. R.; Puchta, H. *Nucleic Acids Res.* **2002**, *30*, 5175–5181.
- (163) Muller, W. E.; Bohm, M.; Grebenjuk, V. A.; Skorokhod, A.; Muller, I. M.; Gamulin, V. *Gene* **2002**, *295*, 299–309.
- (164) Paquette, S. M.; Bak, S.; Feyereisen, R. *DNA Cell Biol.* **2000**, *19*, 307–317.
- (165) Robertson, H. M. *Genome Res.* **1998**, *8*, 449–463.
- (166) Rzhetsky, A.; Ayala, F. J.; Hsu, L. C.; Chang, C.; Yoshida, A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6820–6825.
- (167) Wang, C.; Tymas, M. A.; Butt, T. M. *J. Mol. Evol.* **2005**, *60*, 238–246.
- (168) Kwiatowski, J.; Krawczyk, M.; Kornacki, M.; Bailey, K.; Ayala, F. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8503–8506.
- (169) Logsdon, J. M., Jr.; Tyshenko, M. G.; Dixon, C.; Jafari, J.; Walker, V. K.; Palmer, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8507–8511.
- (170) Straus, D.; Gilbert, W. *Mol. Cell. Biol.* **1985**, *5*, 3497–3506.
- (171) Jeffares, D. C.; Mourier, T.; Penny, D. *Trends Genet.* **2006**, *22*, 16–22.
- (172) Nguyen, H. D.; Yoshihama, M.; Kenmochi, N. *PLoS Comput. Biol.* **2005**, *1*, e79.
- (173) Rogozin, I. B.; Wolf, Y. I.; Sorokin, A. V.; Mirkin, B. G.; Koonin, E. V. *Curr. Biol.* **2003**, *13*, 1512–1517.
- (174) Roy, S. W.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5773–5778.
- (175) Sverdlov, A. V.; Rogozin, I. B.; Babenko, V. N.; Koonin, E. V. *Nucleic Acids Res.* **2005**, *33*, 1741–1748.
- (176) Coulombe-Huntington, J.; Majewski, J. *Genome Res.* **2007**, *17*, 23–32.
- (177) Zhang, X. H.; Chasin, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13427–13432.
- (178) Alekseyenko, A. V.; Kim, N.; Lee, C. J. *RNA* **2007**, *13*, 661–670.
- (179) Couzin, J. *Science* **2003**, *299*, 1642.
- (180) ENCODE Project Consortium. *Science* **2004**, *306*, 636–640.
- (181) Dujon, B. *Trends Genet.* **2006**, *22*, 375–387.
- (182) Bergman, C. M.; Pfeiffer, B. D.; Rincon-Limas, D. E.; Hoskins, R. A.; Gnirke, A.; Mungall, C. J.; Wang, A. M.; Kronmiller, B.; Pacleb, J.; Park, S.; Stapleton, M.; Wan, K.; George, R. A.; de Jong, P. J.; Botas, J.; Rubin, G. M.; Celniker, S. E. *Genome Biol.* **2002**, *3*, research0086.
- (183) Bon, E.; Casaregola, S.; Blandin, G.; Llorente, B.; Neuveglise, C.; Munsterkotter, M.; Guldener, U.; Mewes, H. W.; Van Helden, J.; Dujon, B.; Gaillardin, C. *Nucleic Acids Res.* **2003**, *31*, 1121–1135.
- (184) Cliften, P. F.; Fulton, R. S.; Wilson, R. K.; Johnston, M. *Genetics* **2006**, *172*, 863–872.
- (185) Nielsen, C. B.; Friedman, B.; Birren, B.; Burge, C. B.; Galagan, J. E. *PLoS Biol.* **2004**, *2*, e422.
- (186) Stajich, J. E.; Dietrich, F. S. *Eukaryotic Cell* **2006**, *5*, 789–793.
- (187) Sakurai, A.; Fujimori, S.; Kochiwa, H.; Kitamura-Abe, S.; Washio, T.; Saito, R.; Carninci, P.; Hayashizaki, Y.; Tomita, M. *Gene* **2002**, *300*, 89–95.
- (188) Roy, S. W.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 713–718.
- (189) Sverdlov, A. V.; Babenko, V. N.; Rogozin, I. B.; Koonin, E. V. *Gene* **2004**, *338*, 85–91.
- (190) Chatterji, S.; Pachter, L. *Genomics* **2007**, *90*, 44–48.
- (191) Babenko, V. N.; Rogozin, I. B.; Mekhedov, S. L.; Koonin, E. V. *Nucleic Acids Res.* **2004**, *32*, 3724–3733.
- (192) Roy, S. W.; Fedorov, A.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7158–7162.
- (193) Teich, R.; Grauvogel, C.; Petersen, J. *Gene* **2007**, *394*, 96–104.
- (194) Knowles, D. G.; McLysaght, A. *Mol. Biol. Evol.* **2006**, *23*, 1548–1557.
- (195) Lin, H.; Zhu, W.; Silva, J. C.; Gu, X.; Buell, C. R. *Genome Biol.* **2006**, *7*, R41.
- (196) Edvardsen, R. B.; Lerat, E.; Maeland, A. D.; Flat, M.; Tewari, R.; Jensen, M. F.; Lehrach, H.; Reinhardt, R.; Seo, H. C.; Chourrout, D. *J. Mol. Evol.* **2004**, *59*, 448–457.
- (197) Cho, S.; Jin, S. W.; Cohen, A.; Ellis, R. E. *Genome Res.* **2004**, *14*, 1207–1220.
- (198) Coghlan, A.; Wolfe, K. H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11362–11367.
- (199) Kent, W. J.; Zahler, A. M. *Genome Res.* **2000**, *10*, 1115–1125.
- (200) Stein, L. D.; Bao, Z.; Blasiar, D.; Blumenthal, T.; Brent, M. R.; Chen, N.; Chinwalla, A.; Clarke, L.; Clee, C.; Coghlan, A.; Coulson, A.; D'Eustachio, P.; Fitch, D. H.; Fulton, L. A.; Fulton, R. E.; Griffiths-Jones, S.; Harris, T. W.; Hillier, L. W.; Kamath, R.; Kuwabara, P. E.; Mardis, E. R.; Marra, M. A.; Miner, T. L.; Minx, P.; Mullikin, J. C.; Plumb, R. W.; Rogers, J.; Schein, J. E.; Sohrmann, M.; Spieth, J.; Stajich, J. E.; Wei, C.; Willey, D.; Wilson, R. K.; Durbin, R.; Waterston, R. H. *PLoS Biol.* **2003**, *1*, E45.
- (201) Roy, S. W.; Hartl, D. L. *Genome Res.* **2006**, *16*, 750–756.
- (202) Vanacova, S.; Yan, W.; Carlton, J. M.; Johnson, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4430–4435.
- (203) Fedorov, A.; Roy, S.; Fedorova, L.; Gilbert, W. *Genome Res.* **2003**, *13*, 2236–2241.
- (204) Rogozin, I. B.; Lyons-Weiler, J.; Koonin, E. V. *Trends Genet.* **2000**, *16*, 430–432.
- (205) Dibb, N. J. *J. Theor. Biol.* **1991**, *151*, 405–416.
- (206) Qiu, W. G.; Schisler, N.; Stoltzfus, A. *Mol. Biol. Evol.* **2004**, *21*, 1252–1263.
- (207) Ruvinsky, A.; Eskesen, S. T.; Eskesen, F. N.; Hurst, L. D. *J. Mol. Evol.* **2005**, *60*, 99–104.
- (208) Long, M.; Rosenberg, C. *Mol. Biol. Evol.* **2000**, *17*, 1789–1796.
- (209) Ruvinsky, A.; Ward, W. J. *Mol. Biol. Evol.* **2006**, *23*, 136–141.
- (210) Roy, S. W.; Lewis, B. P.; Fedorov, A.; Gilbert, W. *Trends Genet.* **2001**, *17*, 496–501.
- (211) Sverdlov, A. V.; Rogozin, I. B.; Babenko, V. N.; Koonin, E. V. *Curr. Biol.* **2003**, *13*, 2170–2174.
- (212) Sverdlov, A. V.; Rogozin, I. B.; Babenko, V. N.; Koonin, E. V. *Curr. Biol.* **2004**, *14*, 1505–1508.
- (213) Long, M.; Deutsch, M. *Mol. Biol. Evol.* **1999**, *16*, 1528–1534.
- (214) Lynch, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6118–6123.
- (215) Carmel, I.; Tal, S.; Vigg, I.; Ast, G. *RNA* **2004**, *10*, 828–840.
- (216) Fedorov, A.; Cao, X.; Saxonov, S.; de Souza, S. J.; Roy, S. W.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13177–13182.
- (217) Fedorov, A.; Roy, S.; Cao, X.; Gilbert, W. *Genome Res.* **2003**, *13*, 1155–1157.
- (218) Roy, S. W.; Nosaka, M.; de Souza, S. J.; Gilbert, W. *Gene* **1999**, *238*, 85–91.
- (219) Liu, M.; Grigoriev, A. *Trends Genet.* **2004**, *20*, 399–403.
- (220) Liu, M.; Walch, H.; Wu, S.; Grigoriev, A. *Nucleic Acids Res.* **2005**, *33*, 95–105.
- (221) Banyai, L.; Patthy, L. *FEBS Lett.* **2004**, *565*, 127–132.
- (222) Biderre, C.; Metenier, G.; Vivares, C. P. *Mol. Biochem. Parasitol.* **1998**, *94*, 283–286.
- (223) Breckenridge, D. G.; Watanabe, Y.; Greenwood, S. J.; Gray, M. W.; Schnare, M. N. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 852–856.
- (224) Collins, L.; Penny, D. *Mol. Biol. Evol.* **2005**, *22*, 1053–1066.
- (225) Fast, N. M.; Doolittle, W. F. *Mol. Biochem. Parasitol.* **1999**, *99*, 275–278.
- (226) Ismaili, N.; Perez-Morga, D.; Walsh, P.; Mayeda, A.; Pays, A.; Tebabi, P.; Krainer, A. R.; Pays, E. *Mol. Biochem. Parasitol.* **1999**, *102*, 103–115.
- (227) Nixon, J. E.; Wang, A.; Morrison, H. G.; McArthur, A. G.; Sogin, M. L.; Loftus, B. J.; Samuelson, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3701–3705.
- (228) Vanacova, S.; Yan, W.; Carlton, J. M.; Johnson, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4430–4435.
- (229) Raible, F.; Tessmar-Raible, K.; Osoegawa, K.; Wincker, P.; Jubin, C.; Balavoine, G.; Ferrier, D.; Benes, V.; de Jong, P.; Weissenbach, J.; Bork, P.; Arendt, D. *Science* **2005**, *310*, 1325–1326.
- (230) Roy, S. W. *Trends Genet.* **2006**, *22*, 468–471.
- (231) Cho, S.; Jin, S. W.; Cohen, A.; Ellis, R. E. *Genome Res.* **2004**, *14*, 1207–1220.
- (232) Kiontke, K.; Gavin, N. P.; Raynes, Y.; Roehrig, C.; Piano, F.; Fitch, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9003–9008.
- (233) Patthy, L. *Gene* **1999**, *238*, 103–114.
- (234) Martin, W.; Koonin, E. V. *Nature* **2006**, *440*, 41–45.

- (235) Modrek, B.; Lee, C. J. *Nat. Genet.* **2003**, *34*, 177–180.
- (236) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczy, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissensbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz, R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G. R.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kasprzyk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F. A.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; de Jong, P.; Catanese, J. J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y. J. *Nature* **2001**, *409*, 860–921.
- (237) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Miklos, G. L. G.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McClintosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandich, A.; Zhu, X. *Science* **2001**, *291*, 1304–1351.
- (238) Waterston, R. H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J. F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; Antonarakis, S. E.; Attwood, J.; Baertsch, R.; Bailey, J.; Barlow, K.; Beck, S.; Berry, E.; Birren, B.; Bloom, T.; Bork, P.; Botcherby, M.; Bray, N.; Brent, M. R.; Brown, D. G.; Brown, S. D.; Bult, C.; Burton, J.; Butler, J.; Campbell, R. D.; Carninci, P.; Cawley, S.; Chiaromonte, F.; Chinwalla, A. T.; Church, D. M.; Clamp, M.; Clee, C.; Collins, F. S.; Cook, L. L.; Copley, R. R.; Coulson, A.; Couronne, O.; Cuff, J.; Curwen, V.; Cutts, T.; Daly, M.; David, R.; Davies, J.; Delehaunty, K. D.; Deri, J.; Dermitzakis, E. T.; Dewey, C.; Dickens, N. J.; Diekhans, M.; Dodge, S.; Dubchak, I.; Dunn, D. M.; Eddy, S. R.; Elnitski, L.; Emes, R. D.; Esvara, P.; Eyras, E.; Felsenfeld, A.; Fewell, G. A.; Flicek, P.; Foley, K.; Frankel, W. N.; Fulton, L. A.; Fulton, R. S.; Furey, T. S.; Gage, D.; Gibbs, R. A.; Glusman, G.; Gnerre, S.; Goldman, N.; Goodstadt, L.; Grafham, D.; Graves, T. A.; Green, E. D.; Gregory, S.; Guigo, R.; Guyer, M.; Hardison, R. C.; Haussler, D.; Hayashizaki, Y.; Hillier, L. W.; Hinrichs, A.; Hlavina, W.; Holzer, T.; Hsu, F.; Hua, A.; Hubbard, T.; Hunt, A.; Jackson, I.; Jaffe, D. B.; Johnson, L. S.; Jones, M.; Jones, T. A.; Joy, A.; Kamal, M.; Karlsson, E. K.; Karolchik, D.; Kasprzyk, A.; Kawai, J.; Keibler, E.; Kells, C.; Kent, W. J.; Kirby, A.; Kolbe, D. L.; Korf, I.; Kucherlapati, R. S.; Kulbokas, E. J., III; Kulp, D.; Landers, T.; Leger, J. P.; Leonard, S.; Letunic, I.; Levine, R.; Li, J.; Li, M.; Lloyd, C.; Lucas, S.; Ma, B.; Maglott, D. R.; Mardis, E. R.; Matthews, L.; Mauceli, E.; Mayer, J. H.; McCarthy, M.; McCombie, W. R.; McLaren, S.; McLay, K.; McPherson, J. D.; Meldrim, J.; Meredith, B.; Mesirov, J. P.; Miller, W.; Miner, T. L.; Mongin, E.; Montgomery, K. T.; Morgan, M.; Mott, R.; Mullikin, J. C.; Muzny, D. M.; Nash, W. E.; Nelson, J. O.; Nhan, M. N.; Nicol, R.; Ning, Z.; Nusbaum, C.; O'Connor, M. J.; Okazaki, Y.; Oliver, K.; Overton-Larty, E.; Pachter, L.; Parra, G.; Pepin, K. H.; Peterson, J.; Pevzner, P.; Plumb, R.; Pohl, C. S.; Poliakov, A.; Ponce, T. C.; Ponting, C. P.; Potter, S.; Quail, M.; Reymond, A.; Roe, B. A.; Roskin, K. M.; Rubin, E. M.; Rust, A. G.; Santos, R.; Sapojnikov, V.; Schultz, B.; Schultz, J.; Schwartz, M. S.; Schwartz, S.; Scott, C.; Seaman, S.; Searle, S.; Sharpe, T.; Sheridan, A.; Shownkeen, R.; Sims, S.; Singer, J. B.; Slater, G.; Smit, A.; Smith, D. R.; Spencer, B.; Stabenau, A.; Stange-Thomann, N.; Sugnet, C.; Suyama, M.; Tesler, G.; Thompson, J.; Torrents, D.; Trevaskis, E.; Tromp, J.; Ucla, C.; Ureta-Vidal, A.; Vinson, J. P.; von Niederhausern, A. C.; Wade, C. M.; Wall, M.; Weber, R. J.; Weiss, R.; Wendl, M. C.; West, A. P.; Wetterstrand, K.; Wheeler, R.; Whelan, S.; Wierzbowski, J.; Willey, D.; Williams, S.; Wilson, R. K.; Winter, E.; Worley, K. C.; Wyman, D.; Yang, S.; Yang, S. P.; Zdobnov, E. M.; Zody, M. C.; Lander, E. S. *Nature* **2002**, *420*, 520–562.
- (239) Nurtdinov, R. N.; Artamonova, I. I.; Mironov, A. A.; Gelfand, M. S. *Hum. Mol. Genet.* **2003**, *12*, 1313–1320.
- (240) Sorek, R.; Shamir, R.; Ast, G. *Trends Genet.* **2004**, *20*, 68–71.
- (241) Sugnet, C. W.; Kent, W. J.; Ares, M., Jr.; Haussler, D. *Pac. Symp. Biocomput.* **2004**, 66–77.
- (242) Thanaraj, T. A.; Clark, F.; Muilu, J. *Nucleic Acids Res.* **2003**, *31*, 2544–2552.
- (243) Mitrovich, Q. M.; Anderson, P. *Genes Dev.* **2000**, *14*, 2173–2184.
- (244) Lejeune, F.; Maquat, L. E. *Curr. Opin. Cell. Biol.* **2005**, *17*, 309–315.
- (245) Labow, B. I.; Souba, W. W.; Abcouwer, S. F. *J. Nutr.* **2001**, *131*, 2467S–2474S.
- (246) Jones, R. B.; Wang, F.; Luo, Y.; Yu, C.; Jin, C.; Suzuki, T.; Kan, M.; McKeehan, W. L. *J. Biol. Chem.* **2001**, *276*, 4158–4167.

- (247) Lewis, B. P.; Green, R. E.; Brenner, S. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 189–192.
- (248) Green, R. E.; Lewis, B. P.; Hillman, R. T.; Blanchette, M.; Lareau, L. F.; Garnett, A. T.; Rio, D. C.; Brenner, S. E. *Bioinformatics* **2003**, *19* (Suppl 1), i118–i121.
- (249) Baek, D.; Green, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12813–12818.
- (250) Nurtidinov, R. N.; Neverov, A. D.; Mironov, A. A.; Gelfand, M. S. Conserved and species-specific alternative splicing in mammalian genomes. Unpublished work, 2006.
- (251) Pan, Q.; Saltzman, A. L.; Kim, Y. K.; Misquitta, C.; Shai, O.; Maquat, L. E.; Frey, B. J.; Blencowe, B. J. *Genes Dev.* **2006**, *20*, 153–158.
- (252) Yeo, G. W.; Van Nostrand, E.; Holste, D.; Poggio, T.; Burge, C. B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2850–2855.
- (253) Chen, F. C.; Chen, C. J.; Ho, J. Y.; Chuang, T. J. *BMC Bioinf.* **2006**, *7*, 136.
- (254) Wang, W.; Zheng, H.; Yang, S.; Yu, H.; Li, J.; Jiang, H.; Su, J.; Yang, L.; Zhang, J.; McDermott, J.; Samudrala, R.; Wang, J.; Yang, H.; Yu, J.; Kristiansen, K.; Wong, G. K.; Wang, J. *Genome Res.* **2005**, *15*, 1258–1264.
- (255) Kent, W. J.; Sugnet, C. W.; Furey, T. S.; Roskin, K. M.; Pringle, T. H.; Zahler, A. M.; Haussler, D. *Genome Res.* **2002**, *12*, 996–1006.
- (256) Karolchik, D.; Baertsch, R.; Diekhans, M.; Furey, T. S.; Hinrichs, A.; Lu, Y. T.; Roskin, K. M.; Schwartz, M.; Sugnet, C. W.; Thomas, D. J.; Weber, R. J.; Haussler, D.; Kent, W. J. *Nucleic Acids Res.* **2003**, *31*, 51–54.
- (257) Pan, Q.; Bakowski, M. A.; Morris, Q.; Zhang, W.; Frey, B. J.; Hughes, T. R.; Blencowe, B. J. *Trends Genet.* **2005**, *21*, 73–77.
- (258) Kan, Z.; Garrett-Engle, P. W.; Johnson, J. M.; Castle, J. C. *Nucleic Acids Res.* **2005**, *33*, 5659–5666.
- (259) Rukov, J. L.; Irimia, M.; Mork, S.; Lund, V. K.; Vinther, J.; Arctander, P. *Mol. Biol. Evol.* **2007**, *24*, 909–917.
- (260) Xing, Y.; Lee, C. J. *PLoS Genet.* **2005**, *1*, e34.
- (261) Magen, A.; Ast, G. *Nucleic Acids Res.* **2005**, *33*, 5574–5582.
- (262) Resch, A.; Xing, Y.; Alekseyenko, A.; Modrek, B.; Lee, C. *Nucleic Acids Res.* **2004**, *32*, 1261–1269.
- (263) Malko, D. B.; Makeev, V. J.; Mironov, A. A.; Gelfand, M. S. *Genome Res.* **2006**, *16*, 505–509.
- (264) Kondrashov, F. A.; Koonin, E. V. *Hum. Mol. Genet.* **2001**, *10*, 2661–2669.
- (265) Letunic, I.; Copley, R. R.; Bork, P. *Hum. Mol. Genet.* **2002**, *11*, 1561–1567.
- (266) Kondrashov, F. A.; Koonin, E. V. *Trends Genet.* **2003**, *19*, 115–119.
- (267) Lev-Maor, G.; Sorek, R.; Shomron, N.; Ast, G. *Science* **2003**, *300*, 1288–1291.
- (268) Sorek, R.; Ast, G.; Graur, D. *Genome Res.* **2002**, *12*, 1060–1067.
- (269) Sorek, R.; Lev-Maor, G.; Reznik, M.; Dagan, T.; Belinky, F.; Graur, D.; Ast, G. *Mol. Cell* **2004**, *14*, 221–231.
- (270) Schmucker, D.; Clemens, J. C.; Shu, H.; Worby, C. A.; Xiao, J.; Muda, M.; Dixon, J. E.; Zipursky, S. L. *Cell* **2000**, *101*, 671–684.
- (271) Chen, B. E.; Kondo, M.; Garnier, A.; Watson, F. L.; Puettmann-Holgado, R.; Lamar, D. R.; Schmucker, D. *Cell* **2006**, *125*, 607–620.
- (272) Graveley, B. R.; Kaur, A.; Gunning, D.; Zipursky, S. L.; Rowen, L.; Clemens, J. C. *RNA* **2004**, *10*, 1499–1506.
- (273) Crayton, M. E., III; Powell, B. C.; Vision, T. J.; Giddings, M. C. *BMC Evol. Biol.* **2006**, *6*, 16.
- (274) Copley, R. R. *Trends Genet.* **2004**, *20*, 171–176.
- (275) Chomez, P.; De Becker, O.; Bertrand, M.; De Plaen, E.; Boon, T.; Lucas, S. *Cancer Res.* **2001**, *61*, 5544–5551.
- (276) Artamonova, I. I.; Gelfand, M. S. *J. Mol. Evol.* **2004**, *59*, 620–631.
- (277) De Plaen, E.; Naerhuyzen, B.; De Smet, C.; Szikora, J. P.; Boon, T. *Genomics* **1997**, *40*, 305–313.
- (278) Zheng, C. L.; Fu, X. D.; Gribskov, M. *RNA* **2005**, *11*, 1777–1787.
- (279) Brown, J. W. *Plant J.* **1996**, *10*, 771–780.
- (280) Krawczak, M.; Thomas, N. S.; Hundrieser, B.; Mort, M.; Wittig, M.; Hampe, J.; Cooper, D. N. *Hum. Mutat.* **2006**.
- (281) Nakai, K.; Sakamoto, H. *Gene* **1994**, *141*, 171–177.
- (282) Roca, X.; Sachidanandam, R.; Krainer, A. R. *Nucleic Acids Res.* **2003**, *31*, 6321–6333.
- (283) Vorechovsky, I. *Nucleic Acids Res.* **2006**, *34*, 4630–4641.
- (284) Hiller, M.; Huse, K.; Szafranski, K.; Jahn, N.; Hampe, J.; Schreiber, S.; Backofen, R.; Platzer, M. *Am. J. Hum. Genet.* **2006**, *78*, 291–302.
- (285) Gotea, V.; Makalowski, W. *Trends Genet.* **2006**, *22*, 260–267.
- (286) Eisenberg, E.; Nemzer, S.; Kinar, Y.; Sorek, R.; Rechavi, G.; Levanon, E. Y. *Trends Genet.* **2005**, *21*, 77–81.
- (287) Athanasiadis, A.; Rich, A.; Maas, S. *PLoS Biol.* **2004**, *2*, e391.
- (288) Kim, D. D.; Kim, T. T.; Walsh, T.; Kobayashi, Y.; Matisse, T. C.; Buyske, S.; Gabriel, A. *Genome Res.* **2004**, *14*, 1719–1725.
- (289) Knebelmann, B.; Forestier, L.; Drouot, L.; Quinones, S.; Chuet, C.; Benessy, F.; Saus, J.; Antignac, C. *Hum. Mol. Genet.* **1995**, *4*, 675–679.
- (290) Mitchell, G. A.; Labuda, D.; Fontaine, G.; Saudubray, J. M.; Bonnefont, J. P.; Lyonnet, S.; Brody, L. C.; Steel, G.; Obie, C.; Valle, D. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 815–819.
- (291) Vervoort, R.; Gitzelmann, R.; Lissens, W.; Liebaers, I. *Hum. Genet.* **1998**, *103*, 686–693.
- (292) Krull, M.; Brosius, J.; Schmitz, J. *Mol. Biol. Evol.* **2005**, *22*, 1702–1711.
- (293) Singer, S. S.; Mannel, D. N.; Hehlgans, T.; Brosius, J.; Schmitz, J. *J. Mol. Biol.* **2004**, *341*, 883–886.
- (294) Mola, G.; Vela, E.; Fernandez-Figueras, M. T.; Isamat, M.; Munoz-Marmol, A. M. *J. Mol. Biol.* **2007**, *366*, 1055–1063.
- (295) Wallace, M. R.; Andersen, L. B.; Saulino, A. M.; Gregory, P. E.; Glover, T. W.; Collins, F. S. *Nature* **1991**, *353*, 864–866.
- (296) Liang, H.; Landweber, L. F. *Genome Res.* **2006**, *16*, 190–196.
- (297) Ciccarelli, F. D.; von Mering, C.; Suyama, M.; Harrington, E. D.; Izaurrealde, E.; Bork, P. *Genome Res.* **2005**, *15*, 343–351.
- (298) Kopelman, N. M.; Lancet, D.; Yanai, I. *Nat. Genet.* **2005**, *37*, 588–589.
- (299) Su, Z.; Wang, J.; Yu, J.; Huang, X.; Gu, X. *Genome Res.* **2006**, *16*, 182–189.
- (300) Brunet, F. G.; Crollius, H. R.; Paris, M.; Aury, J. M.; Gibert, P.; Jaillon, O.; Laudet, V.; Robinson-Rechavi, M. *Mol. Biol. Evol.* **2006**, *23*, 1808–1816.
- (301) Christoffels, A.; Koh, E. G.; Chia, J. M.; Brenner, S.; Aparicio, S.; Venkatesh, B. *Mol. Biol. Evol.* **2004**, *21*, 1146–1151.
- (302) Jaillon, O.; Aury, J. M.; Brunet, F.; Petit, J. L.; Stange-Thomann, N.; Mauceli, E.; Bouneau, L.; Fischer, C.; Ozouf-Costaz, C.; Bernot, A.; Nicaud, S.; Jaffe, D.; Fisher, S.; Lutfalla, G.; Dossat, C.; Segurens, B.; Dasilva, C.; Salanoubat, M.; Levy, M.; Boudet, N.; Castellano, S.; Anthonard, V.; Jubin, C.; Castelli, V.; Katinka, M.; Vacherie, B.; Biemont, C.; Skalli, Z.; Cattolico, L.; Poulain, J.; de Berardinis, V.; Cruaud, C.; Duprat, S.; Brottier, P.; Coutanceau, J. P.; Gouzy, J.; Parra, G.; Lardier, G.; Chapple, C.; McKernan, K. J.; McEwan, P.; Bosak, S.; Kellis, M.; Volff, J. N.; Guigo, R.; Zody, M. C.; Mesirov, J.; Lindblad-Toh, K.; Birren, B.; Nusbaum, C.; Kahn, D.; Robinson-Rechavi, M.; Laudet, V.; Schachter, V.; Quetier, F.; Saurin, W.; Scarpelli, C.; Wincker, P.; Lander, E. S.; Weissenbach, J.; Roest, C. H. *Nature* **2004**, *431*, 946–957.
- (303) Postlethwait, J.; Amores, A.; Cresko, W.; Singer, A.; Yan, Y. L. *Trends Genet.* **2004**, *20*, 481–490.
- (304) Talavera, D.; Vogel, C.; Orozco, M.; Teichmann, S. A.; de la Cruz, X. *PLoS Comput. Biol.* **2007**, *3*, e33.
- (305) Jordan, I. K.; Wolf, Y. I.; Koonin, E. V. *BMC Evol. Biol.* **2004**, *4*, 22.
- (306) Lynch, M.; Conery, J. S. *Science* **2000**, *290*, 1151–1155.
- (307) Nembaware, V.; Crum, K.; Kelso, J.; Seoighe, C. *Genome Res.* **2002**, *12*, 1370–1376.
- (308) van de Peer, Y.; Taylor, J. S.; Braasch, I.; Meyer, A. *J. Mol. Evol.* **2001**, *53*, 436–446.
- (309) Davis, J. C.; Petrov, D. A. *PLoS Biol.* **2004**, *2*, E55.
- (310) Iida, K.; Akashi, H. *Gene* **2000**, *261*, 93–105.
- (311) Xing, Y.; Lee, C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13526–13531.
- (312) Chen, F. C.; Wang, S. S.; Chen, C. J.; Li, W. H.; Chuang, T. J. *Mol. Biol. Evol.* **2006**, *23*, 675–682.
- (313) Ermakova, E. O.; Nurtidinov, R. N.; Gelfand, M. S. *BMC Genomics* **2006**, *7*, 84.
- (314) Ermakova, E. O.; Malko, D. B.; Gelfand, M. S. *Biophysics* **2006**, *51*, 515–522.
- (315) Parmley, J. L.; Chamary, J. V.; Hurst, L. D. *Mol. Biol. Evol.* **2006**, *23*, 301–309.
- (316) Parmley, J. L.; Hurst, L. D. *Mol. Biol. Evol.* **2007**.
- (317) Parmley, J. L.; Urrutia, A. O.; Potrzebowski, L.; Kaessmann, H.; Hurst, L. D. *PLoS Biol.* **2007**, *5*, e14.
- (318) Feng-Chi, C.; Shu-Miaw, C.; Yun-Huei, T.; Sheng-Shun, W.; Trees-Juen, C. *Mol. Biol. Evol.* **2007**.
- (319) Cusack, B. P.; Wolfe, K. H. *Mol. Biol. Evol.* **2005**, *22*, 2198–2208.
- (320) Plass, M.; Eyraes, E. *BMC Evol. Biol.* **2006**, *6*, 50.
- (321) Dewey, C. N.; Rogozin, I. B.; Koonin, E. V. *BMC Genomics* **2006**, *7*, 311.
- (322) Chen, F. C.; Chuang, T. J. *BMC Bioinf.* **2006**, *7*, 259.
- (323) Kondrashov, F. A.; Ogurtsov, A. Y.; Kondrashov, A. S. *J. Theor. Biol.* **2006**, *240*, 616–626.
- (324) McDonald, J. H.; Kreitman, M. *Nature* **1991**, *351*, 652–654.
- (325) Ramensky, V. E.; Nurtidinov, R. N.; Neverov, A. D.; Mironov, A. A.; Gelfand, M. S. Human genome polymorphism and alternative splicing. Proceedings of the 5th International Conference on "Bioinformatics of Genome Regulation and Structure" BGRS'2006; Siberian Branch of the Russian Academy of Sciences: Novosibirsk, Russia, 2006, pp 211–213.

- (326) Itoh, H.; Washio, T.; Tomita, M. *RNA* **2004**, *10*, 1005–1018.
- (327) Xing, Y.; Wang, Q.; Lee, C. *Genetics* **2006**, *173*, 1787–1791.
- (328) Fairbrother, W. G.; Holste, D.; Burge, C. B.; Sharp, P. A. *PLoS Biol.* **2004**, *2*, E268.
- (329) Schattner, P.; Diekhans, M. *Nucleic Acids Res.* **2006**, *34*, 1700–1710.
- (330) Carlini, D. B.; Genut, J. E. *J. Mol. Evol.* **2006**, *62*, 89–98.
- (331) Pozzoli, U.; Sironi, M. *Cell. Mol. Life Sci.* **2005**, *62*, 1579–1604.
- (332) Wang, Z.; Rolish, M. E.; Yeo, G.; Tung, V.; Mawson, M.; Burge, C. B. *Cell* **2004**, *119*, 831–845.
- (333) Blanchette, M. Personal communication, 2006
- (334) Philipps, D. L.; Park, J. W.; Graveley, B. R. *RNA* **2004**, *10*, 1838–1844.
- (335) Sorek, R.; Ast, G. *Genome Res.* **2003**, *13*, 1631–1637.
- (336) Sugnet, C. W.; Srinivasan, A.; Clark, T. A.; O'Brien, G.; Cline, M. S.; Wang, H.; Williams, A.; Kulp, D.; Blume, J. E.; Haussler, D.; Ares, M., Jr. *PLoS Comput. Biol.* **2006**, *2*, e4.
- (337) Vinogradov, A. E. *Genome Res.* **2006**, *16*, 347–354.
- (338) Wang, Z.; Xiao, X.; Van Nostrand, E.; Burge, C. B. *Mol. Cell* **2006**, *23*, 61–70.
- (339) Brudno, M.; Gelfand, M. S.; Spengler, S.; Zorn, M.; Dubchak, I.; Conboy, J. G. *Nucleic Acids Res.* **2001**, *29*, 2338–2348.
- (340) Chamary, J. V.; Hurst, L. D. *Mol. Biol. Evol.* **2004**, *21*, 1014–1023.
- (341) Akerman, M.; Mandel-Gutfreund, Y. *Nucleic Acids Res.* **2006**, *34*, 23–31.
- (342) Hiller, M.; Huse, K.; Szafranski, K.; Rosenstiel, P.; Schreiber, S.; Backofen, R.; Platzer, M. *Genome Biol.* **2006**, *7*, R65.
- (343) Yeo, G.; Hoon, S.; Venkatesh, B.; Burge, C. B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15700–15705.
- (344) Jin, Y.; Suzuki, H.; Maegawa, S.; Endo, H.; Sugano, S.; Hashimoto, K.; Yasuda, K.; Inoue, K. *EMBO J.* **2003**, *22*, 905–912.
- (345) Minovitsky, S.; Gee, S. L.; Schokrpur, S.; Dubchak, I.; Conboy, J. G. *Nucleic Acids Res.* **2005**, *33*, 714–724.
- (346) Kabat, J. L.; Barberan-Soler, S.; McKenna, P.; Clawson, H.; Farrer, T.; Zahler, A. M. *PLoS Comput. Biol.* **2006**, *2*, e86.
- (347) Nicoll, M.; Akerib, C. C.; Meyer, B. J. *Nature* **1997**, *388*, 200–204.
- (348) Andolfatto, P. *Nature* **2005**, *437*, 1149–1152.
- (349) Fox-Walsh, K. L.; Dou, Y.; Lam, B. J.; Hung, S. P.; Baldi, P. F.; Hertel, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16176–16181.
- (350) Burnette, J. M.; Miyamoto-Sato, E.; Schaub, M. A.; Conklin, J.; Lopez, A. J. *Genetics* **2005**, *170*, 661–674.
- (351) Sorek, R.; Shemesh, R.; Cohen, Y.; Basechess, O.; Ast, G.; Shamir, R. *Genome Res.* **2004**, *14*, 1617–1623.
- (352) Dror, G.; Sorek, R.; Shamir, R. *Bioinformatics* **2005**, *21*, 897–901.
- (353) Ohler, U.; Shomron, N.; Burge, C. B. *PLoS Comput. Biol.* **2005**, *1*, 113–122.
- (354) Chuang, T. J.; Chen, F. C.; Chou, M. Y. *Bioinformatics* **2004**, *20*, 3064–3079.
- (355) Vukusic, I.; Grellscheid, S. N.; Wiehe, T. *Genomics* **2007**, *89*, 471–479.
- (356) Hiller, M.; Huse, K.; Platzer, M.; Backofen, R. *Nucleic Acids Res.* **2005**, *33*, 5611–5621.
- (357) Ratsch, G.; Sonnenburg, S.; Scholkopf, B. *Bioinformatics* **2005**, *21 Suppl 1*, i369–i377.
- (358) Fairbrother, W. G.; Yeh, R. F.; Sharp, P. A.; Burge, C. B. *Science* **2002**, *297*, 1007–1013.
- (359) Fairbrother, W. G.; Yeo, G. W.; Yeh, R.; Goldstein, P.; Mawson, M.; Sharp, P. A.; Burge, C. B. *Nucleic Acids Res.* **2004**, *32*, W187–W190.
- (360) Cartegni, L.; Wang, J.; Zhu, Z.; Zhang, M. Q.; Krainer, A. R. *Nucleic Acids Res.* **2003**, *31*, 3568–3571.
- (361) Wang, J.; Smith, P. J.; Krainer, A. R.; Zhang, M. Q. *Nucleic Acids Res.* **2005**, *33*, 5053–5062.
- (362) Zhang, X. H.; Chasin, L. A. *Genes Dev.* **2004**, *18*, 1241–1250.
- (363) Sironi, M.; Menozzi, G.; Riva, L.; Cagliani, R.; Comi, G. P.; Bresolin, N.; Giorda, R.; Pozzoli, U. *Nucleic Acids Res.* **2004**, *32*, 1783–1791.
- (364) Hanke, J.; Brett, D.; Zastrow, I.; Aydin, A.; Delbruck, S.; Lehmann, G.; Luft, F.; Reich, J.; Bork, P. *Trends Genet.* **1999**, *15*, 389–390.
- (365) Croft, L.; Schandorff, S.; Clark, F.; Burrage, K.; Arctander, P.; Mattick, J. S. *Nat. Genet.* **2000**, *24*, 340–341.
- (366) Kan, Z.; Rouchka, E. C.; Gish, W. R.; States, D. J. *Genome Res.* **2001**, *11*, 889–900.
- (367) Deloukas, P.; Matthews, L. H.; Ashurst, J.; Burton, J.; Gilbert, J. G. R.; Jones, M.; Stavrides, G.; Almeida, J. P.; Babbage, A. K.; Bagguley, C. L.; Bailey, J.; Barlow, K. F.; Bates, K. N.; Beard, L. M.; Beare, D. M.; Beasley, O. P.; Bird, C. P.; Blakey, S. E.; Bridgeman, A. M.; Brown, A. J.; Buck, D.; Burrill, W.; Butler, A. P.; Carder, C.; Carter, N. P.; Chapman, J. C.; Clamp, M.; Clark, G.; Clark, L. N.; Clark, S. Y.; Clee, C. M.; Clegg, S.; Copley, V. E.; Collier, R. E.; Connor, R.; Corby, N. R.; Coulson, A.; Coville, G. J.; Deadman, R.; Dhami, P.; Dunn, M.; Ellington, A. G.; Frankland, J. A.; Fraser, A.; French, L.; Garner, P.; Grafham, D. V.; Griffiths, C.; Griffiths, M. N. D.; Gilliam, R.; Hall, R. E.; Hammond, S.; Harley, J. L.; Heath, P. D.; Ho, S.; Holden, J. L.; Howden, P. J.; Huckle, E.; Hunt, A. R.; Hunt, S. E.; Jekosch, K.; Johnson, C. M.; Johnson, D.; Kay, M. P.; Kimberley, A. M.; King, A.; Knights, A.; Laird, G. K.; Lawlor, S.; Lehtvaslahti, M. H.; Leversha, M.; Lloyd, C.; Lloyd, D. M.; Lovell, J. D.; Marsh, V. L.; Martin, S. L.; McConnachie, L. J.; McLay, K.; McMurray, A. A.; Milne, S.; Mistry, D.; Moore, M. J. F.; Mullikin, J. C.; Nickerson, T.; Oliver, K.; Parker, A.; Patel, R.; Pearce, T. A. V.; Peck, A. I.; Phillimore, B. J. C. T.; Prathalingam, S. R.; Plumb, R. W.; Ramsay, H.; Rice, C. M.; Ross, M. T.; Scott, C. E.; Sehra, H. K.; Showkeen, R.; Sims, S.; Skuce, C. D.; Smith, M. L.; Soderlund, C.; Steward, C. A.; Sulston, J. E.; Swann, M.; Sycamore, N.; Taylor, R.; Tee, L.; Thomas, D. W.; Thorpe, A.; Tracey, A.; Tromans, A. C.; Vaudin, M.; Wall, M.; Wallis, J. M.; Whitehead, S. L.; Whittaker, P.; Willey, D. L.; Williams, L.; Williams, S. A.; Wilming, L.; Wray, P. W.; Hubbard, T.; Durbin, R. M.; Bentley, D. R.; Beck, S.; Rogers, J. *Nature* **2001**, *414*, 865–871.
- (368) Clark, F.; Thanaraj, T. A. *Hum. Mol. Genet.* **2002**, *11*, 451–464.
- (369) Zavolan, M.; van Nimwegen, E.; Gaasterland, T. *Genome Res.* **2002**, *12*, 1377–1385.
- (370) Okazaki, Y.; Furuno, M.; Kasukawa, T.; Adachi, J.; Bono, H.; Kondo, S.; Nikaïdo, I.; Osato, N.; Saito, R.; Suzuki, H.; Yamanaka, I.; Kiyosawa, H.; Yagi, K.; Tomaru, Y.; Hasegawa, Y.; Nogami, A.; Schonbach, C.; Gojbori, T.; Baldarelli, R.; Hill, D. P.; Bult, C.; Hume, D. A.; Quackenbush, J.; Schriml, L. M.; Kanapin, A.; Matsuda, H.; Batalov, S.; Beisel, K. W.; Blake, J. A.; Bradt, D.; Brusci, V.; Chothia, C.; Corbani, L. E.; Cousins, S.; Dalla, E.; Dragani, T. A.; Fletcher, C. F.; Forrest, A.; Frazer, K. S.; Gaasterland, T.; Gariboldi, M.; Gissi, C.; Godzik, A.; Gough, J.; Grimmond, S.; Gustinich, S.; Hirokawa, N.; Jackson, I. J.; Jarvis, E. D.; Kanai, A.; Kawaji, H.; Kawasawa, Y.; Kedzierski, R. M.; King, B. L.; Konagaya, A.; Kurochkin, I. V.; Lee, Y.; Lenhard, B.; Lyons, P. A.; Maglott, D. R.; Maltais, L.; Marchionni, L.; McKenzie, L.; Miki, H.; Nagashima, T.; Numata, K.; Okido, T.; Pavan, W. J.; Perlea, G.; Pesole, G.; Petrovsky, N.; Pillai, R.; Pontius, J. U.; Qi, D.; Ramachandran, S.; Ravasi, T.; Reed, J. C.; Reed, D. J.; Reid, J.; Ring, B. Z.; Ringwald, M.; Sandelin, A.; Schneider, C.; Sempile, C. A. M.; Setou, M.; Shimada, K.; Sultana, R.; Takenaka, Y.; Taylor, M. S.; Teasdale, R. D.; Tomita, M.; Verardo, R.; Wagner, L.; Wahlestedt, C.; Wang, Y.; Watanabe, Y.; Wells, C.; Wilming, L. G.; Wynshaw-Boris, A.; Yanagisawa, M.; Yang, I.; Yang, L.; Yuan, Z.; Zavolan, M.; Zhu, Y.; Zimmer, A.; Carninci, P.; Hayatsu, N.; Hirozane-Kishikawa, T.; Konno, H.; Nakamura, M.; Sakazume, N.; Sato, K.; Shiraki, T.; Waki, K.; Kawai, J.; Aizawa, K.; Arakawa, T.; Fukuda, S.; Hara, A.; Hashizume, W.; Imotani, K.; Ishii, Y.; Itoh, M.; Kagawa, I.; Miyazaki, A.; Sakai, K.; Sasaki, D.; Shibata, K.; Shinagawa, A.; Yasunishi, A.; Yoshino, M.; Waterston, R.; Lander, E. S.; Rogers, J.; Birney, E.; Hayashizaki, Y. *Nature* **2002**, *420*, 563–573.
- (371) Heilig, R.; Eckenberg, R.; Petit, J. L.; Fonknechten, N.; Da Silva, C.; Cattoico, L.; Levy, M.; Barbe, V.; de Berardinis, V.; Ureta-Vidal, A.; Pelletier, E.; Vico, V.; Anthouard, V.; Rowen, L.; Madan, A.; Qin, S.; Sun, H.; Du, H.; Pepin, K.; Artiguenave, F.; Robert, C.; Cruaud, C.; Bruls, T.; Jaillon, O.; Friedlander, L.; Samson, G.; Brottier, P.; Cure, S.; Segurens, B.; Aniere, F.; Samain, S.; Crespeau, H.; Abbasi, N.; Aiach, N.; Boscus, D.; Dickhoff, R.; Dors, M.; Dubois, I.; Friedman, C.; Gouyvenoux, M.; James, R.; Madan, A.; Mairey-Estrada, B.; Mangenot, S.; Martins, N.; Menard, M.; Oztas, S.; Ratcliffe, A.; Shaffer, T.; Trask, B.; Vacherie, B.; Bellemere, C.; Belser, C.; Besnard-Gonnet, M.; Bartol-Mavel, D.; Boutard, M.; Briez-Silla, S.; Combette, S.; Dufosse-Laurent, V.; Ferron, C.; Lechaplais, C.; Louesse, C.; Muselet, D.; Magdelenat, G.; Pateau, E.; Petit, E.; Sirvain-Trukniewicz, P.; Trybou, A.; Vega-Czarny, N.; Bataille, E.; Bluet, E.; Bordelais, I.; Dubois, M.; Dumont, C.; Guerin, T.; Haffray, S.; Hammadi, R.; Muanga, J.; Pellouin, V.; Robert, D.; Wunderle, E.; Gauguier, G.; Roy, A.; Sainte-Marthe, L.; Verdier, J.; Verdier-Discalca, C.; Hillier, L.; Fulton, L.; McPherson, J.; Matsuda, F.; Wilson, R.; Scarpelli, C.; Gyapay, G.; Wincker, P.; Saurin, W.; Quetier, F.; Waterston, R.; Hood, L.; Weissbach, J. *Nature* **2003**, *421*, 601–607.
- (372) Zavolan, M.; Kondo, S.; Schonbach, C.; Adachi, J.; Hume, D. A.; Hayashizaki, Y.; Gaasterland, T. *Genome Res.* **2003**, *13*, 1290–1300.
- (373) Hillier, L. W.; Fulton, R. S.; Fulton, L. A.; Graves, T. A.; Pepin, K. H.; Wagner-McPherson, C.; Layman, D.; Maas, J.; Jaeger, S.; Walker, R.; Wylie, K.; Sekhon, M.; Becker, M. C.; O'Laughlin, M. D.; Schaller, M. E.; Fewell, G. A.; Delehaunty, K. D.; Miner, T. L.; Nash, W. E.; Cordes, M.; Du, H.; Sun, H.; Edwards, J.; Bradshaw-Cordum, H.; Ali, J.; Andrews, S.; Isak, A.; VanBrunt, A.; Nguyen, C.; Du, F.; Lamar, B.; Courtney, L.; Kalicki, J.; Ozersky, P.; Bielicki, L.; Scott, K.; Holmes, A.; Harkins, R.; Harris, A.; Strong, C. M.; Hou, S.; Tomlinson, C.; Dauphin-Kohlberg, S.; Kozlowski-Reilly, A.; Leonard, S.; Rohlfling, T.; Rock, S. M.; Tin-Wollam, A. M.; Abbott, A.; Minx, P.; Maupin, R.; Strowmatt, C.; Latreille, P.; Miller, N.; Johnson, D.; Murray, J.; Woessner, J. P.; Wendl, M. C.; Yang,

- S. P.; Schultz, B. R.; Wallis, J. W.; Spieth, J.; Bieri, T. A.; Nelson, J. O.; Berkowicz, N.; Wohldmann, P. E.; Cook, L. L.; Hickenbotham, M. T.; Eldred, J.; Williams, D.; Bedell, J. A.; Mardis, E. R.; Clifton, S. W.; Chissoe, S. L.; Marra, M. A.; Raymond, C.; Haugen, E.; Gillett, W.; Zhou, Y.; James, R.; Phelps, K.; Iadanoto, S.; Bubb, K.; Simms, E.; Levy, R.; Clendenning, J.; Kaul, R.; Kent, W. J.; Furey, T. S.; Baertsch, R. A.; Brent, M. R.; Keibler, E.; Flicek, P.; Bork, P.; Suyama, M.; Bailey, J. A.; Portnoy, M. E.; Torrents, D.; Chinwalla, A. T.; Gish, W. R.; Eddy, S. R.; McPherson, J. D.; Olson, M. V.; Eichler, E. E.; Green, E. D.; Waterston, R. H.; Wilson, R. K. *Nature* **2003**, *424*, 157–164.
- (374) Grimwood, J.; Gordon, L. A.; Olsen, A.; Terry, A.; Schmutz, J.; Lamerdin, J.; Hellsten, U.; Goodstein, D.; Couronne, O.; Tran-Gyamfi, M.; Aerts, A.; Altherr, M.; Ashworth, L.; Bajorek, E.; Black, S.; Branscomb, E.; Caenepeel, S.; Carrano, A.; Caoile, C.; Chan, Y. M.; Christensen, M.; Cleland, C. A.; Copeland, A.; Dalin, E.; Dehal, P.; Denys, M.; Detter, J. C.; Escobar, J.; Flowers, D.; Fotopulos, D.; Garcia, C.; Georgescu, A. M.; Glavina, T.; Gomez, M.; Gonzales, E.; Groza, M.; Hammon, N.; Hawkins, T.; Haydu, L.; Ho, I.; Huang, W.; Israni, S.; Jett, J.; Kadner, K.; Kimball, H.; Kobayashi, A.; Larionov, V.; Leem, S. H.; Lopez, F.; Lou, Y.; Lowry, S.; Malfatti, S.; Martinez, D.; McCready, P.; Medina, C.; Morgan, J.; Nelson, K.; Nolan, M.; Ovcharenko, I.; Pitluck, S.; Pollard, M.; Popkie, A. P.; Predki, P.; Quan, G.; Ramirez, L.; Rash, S.; Retterer, J.; Rodriguez, A.; Rogers, S.; Salamov, A.; Salazar, A.; She, X.; Smith, D.; Slezak, T.; Solovyev, V.; Thayer, N.; Tice, H.; Tsai, M.; Ustaszewska, A.; Vo, N.; Wagner, M.; Wheeler, J.; Wu, K.; Xie, G.; Yang, J.; Dubchak, I.; Furey, T. S.; DeJong, P.; Dickson, M.; Gordon, D.; Eichler, E. E.; Pennacchio, L. A.; Richardson, P.; Stubbs, L.; Rokhsar, D. S.; Myers, R. M.; Rubin, E. M.; Lucas, S. M. *Nature* **2004**, *428*, 529–535.
- (375) Deloukas, P.; Earthrowl, M. E.; Grafham, D. V.; Rubenfield, M.; French, L.; Steward, C. A.; Sims, S. K.; Jones, M. C.; Searle, S.; Scott, C.; Howe, K.; Hunt, S. E.; Andrews, T. D.; Gilbert, J. G. R.; Swarbreck, D.; Ashurst, J. L.; Taylor, A.; Battles, J.; Bird, C. P.; Ainscough, R.; Almeida, J. P.; Ashwell, R. I. S.; Ambrose, K. D.; Babbage, A. K.; Bagguley, C. L.; Bailey, J.; Banerjee, R.; Bates, K.; Beasley, H.; Bray-Allen, S.; Brown, A. J.; Brown, J. Y.; Burford, D. C.; Burrill, W.; Burton, J.; Cahill, P.; Camire, D.; Carter, N. P.; Chapman, J. C.; Clark, S. Y.; Clarke, G.; Clee, C. M.; Clegg, S.; Corby, N.; Coulson, A.; Dhami, P.; Dutta, I.; Dunn, M.; Faulkner, L.; Frankish, A.; Frankland, J. A.; Garner, P.; Garnett, J.; Gribble, S.; Griffiths, C.; Grocock, R.; Gustafson, E.; Hammond, S.; Harley, J. L.; Hart, E.; Heath, P. D.; Ho, T. P.; Hopkins, B.; Horne, J.; Howden, P. J.; Huckle, E.; Hynds, C.; Johnson, C.; Johnson, D.; Kana, A.; Kay, M.; Kimberley, A. M.; Kershaw, J. K.; Kokkinaki, M.; Laird, G. K.; Lawlor, S.; Lee, H. M.; Leongamornlert, D. A.; Laird, G.; Lloyd, C.; Lloyd, D. M.; Loveland, J.; Lovell, J.; McLaren, S.; McLay, K. E.; McMurray, A.; Mashreghi-Mohammadi, M.; Matthews, L.; Milne, S.; Nickerson, T.; Nguyen, M.; Overton-Larty, E.; Palmer, S. A.; Pearce, A. V.; Peck, A. I.; Pelan, S.; Phillimore, B.; Porter, K.; Rice, C. M.; Rogosin, A.; Ross, M. T.; Sarafidou, T.; Sehra, H. K.; Shownkeen, R.; Skuce, C. D.; Smith, M.; Standring, L.; Sycamore, N.; Tester, J.; Thorpe, A.; Torcasso, W.; Tracey, A.; Tromans, A.; Tsolas, J.; Wall, M.; Walsh, J.; Wang, H.; Weinstock, K.; West, A. P.; Willey, D. L.; Whitehead, S. L.; Wilming, L.; Wray, P. W.; Young, L.; Chen, Y.; Lovering, R. C.; Moschonas, N. K.; Siebert, R.; Fechtel, K.; Bentley, D.; Durbin, R.; Hubbard, T.; Doucette-Stamm, L.; Beck, S.; Smith, D. R.; Rogers, J. *Nature* **2004**, *429*, 375–381.
- (376) Imanishi, T.; Itoh, T.; Suzuki, Y.; O'Donovan, C.; Fukuchi, S.; Koyanagi, K. O.; Barrero, R. A.; Tamura, T.; Yamaguchi-Kabata, Y.; Tanino, M.; Yura, K.; Miyazaki, S.; Ikeo, K.; Homma, K.; Kasprzyk, A.; Nishikawa, T.; Hirakawa, M.; Thierry-Mieg, J.; Thierry-Mieg, D.; Ashurst, J.; Jia, L.; Nakao, M.; Thomas, M. A.; Mulder, N.; Karavidopoulou, Y.; Jin, L.; Kim, S.; Yasuda, T.; Lenhard, B.; Eveno, E.; Suzuki, Y.; Yamasaki, C.; Takeda, J.; Gough, C.; Hilton, P.; Fujii, Y.; Sakai, H.; Tanaka, S.; Amid, C.; Bellgard, M.; Bonaldo, M. F.; Bono, H.; Bromberg, S. K.; Brookes, A. J.; Bruford, E.; Carninci, P.; Chelala, C.; Couillault, C.; de Souza, S. J.; Debily, M. A.; Devignes, M. D.; Dubchak, I.; Endo, T.; Estreicher, A.; Eyra, E.; Fukami-Kobayashi, K.; Gopinath, G. R.; Graudens, E.; Hahn, Y.; Han, M.; Han, Z. G.; Hanada, K.; Hanaoka, H.; Harada, E.; Hashimoto, K.; Hinz, U.; Hirai, M.; Hishiki, T.; Hopkinson, I.; Imbeaud, S.; Inoko, H.; Kanapin, A.; Kaneko, Y.; Kasukawa, T.; Kelso, J.; Kersey, P.; Kikuno, R.; Kimura, K.; Korn, B.; Kuryshhev, V.; Makalowska, I.; Makino, T.; Mano, S.; Mariage-Samson, R.; Mashima, J.; Matsuda, H.; Mewes, H. W.; Minoshima, S.; Nagai, K.; Nagasaki, H.; Nagata, N.; Nigam, R.; Ogasawara, O.; Ohara, O.; Ohtsubo, M.; Okada, N.; Okido, T.; Oota, S.; Ota, M.; Ota, T.; Otsuki, T.; Piatier-Tonneau, D.; Poustka, A.; Ren, S. X.; Saitou, N.; Sakai, K.; Sakamoto, S.; Sakate, R.; Schupp, I.; Servant, F.; Sherry, S.; Shiba, R.; Shimizu, N.; Shimoyama, M.; Simpson, A. J.; Soares, B.; Steward, C.; Suwa, M.; Suzuki, M.; Takahashi, A.; Tamiya, G.; Tanaka, H.; Taylor, T.; Terwilliger, J. D.; Unneberg, P.; Veeramachaneni, V.; Watanabe, S.; Wilming, L.; Yasuda, N.; Yoo, H. S.; Stodolsky, M.; Makalowski, W.; Go, M.; Nakai, K.; Takagi, T.; Kanehisa, M.; Sakaki, Y.; Quackenbush, J.; Okazaki, Y.; Hayashizaki, Y.; Hide, W.; Chakraborty, R.; Nishikawa, K.; Sugawara, H.; Tateno, Y.; Chen, Z.; Oishi, M.; Tonellato, P.; Apweiler, R.; Okubo, K.; Wagner, L.; Wiemann, S.; Strausberg, R. L.; Isogai, T.; Auffray, C.; Nomura, N.; Gojobori, T.; Sugano, S. *PLoS Biol.* **2004**, *2*, e162.
- (377) Schmutz, J.; Martin, J.; Terry, A.; Couronne, O.; Grimwood, J.; Lowry, S.; Gordon, L. A.; Scott, D.; Xie, G.; Huang, W.; Hellsten, U.; Tran-Gyamfi, M.; She, X.; Prabhakar, S.; Aerts, A.; Altherr, M.; Bajorek, E.; Black, S.; Branscomb, E.; Caoile, C.; Challacombe, J. F.; Chan, Y. M.; Denys, M.; Detter, J. C.; Escobar, J.; Flowers, D.; Fotopulos, D.; Glavina, T.; Gomez, M.; Gonzales, E.; Goodstein, D.; Grigoriev, I.; Groza, M.; Hammon, N.; Hawkins, T.; Haydu, L.; Israni, S.; Jett, J.; Kadner, K.; Kimball, H.; Kobayashi, A.; Lopez, F.; Lou, Y.; Martinez, D.; Medina, C.; Morgan, J.; Nandkeshwar, R.; Noonan, J. P.; Pitluck, S.; Pollard, M.; Predki, P.; Priest, J.; Ramirez, L.; Retterer, J.; Rodriguez, A.; Rogers, S.; Salamov, A.; Salazar, A.; Thayer, N.; Tice, H.; Tsai, M.; Ustaszewska, A.; Vo, N.; Wheeler, J.; Wu, K.; Yang, J.; Dickson, M.; Cheng, J. F.; Eichler, E. E.; Olsen, A.; Pennacchio, L. A.; Rokhsar, D. S.; Richardson, P.; Lucas, S. M.; Myers, R. M.; Rubin, E. M. *Nature* **2004**, *431*, 268–274.
- (378) Martin, J.; Han, C.; Gordon, L. A.; Terry, A.; Prabhakar, S.; She, X.; Xie, G.; Hellsten, U.; Chan, Y. M.; Altherr, M.; Couronne, O.; Aerts, A.; Bajorek, E.; Black, S.; Blumer, H.; Branscomb, E.; Brown, N. C.; Bruno, W. J.; Buckingham, J. M.; Callen, D. F.; Campbell, C. S.; Campbell, M. L.; Campbell, E. W.; Caoile, C.; Challacombe, J. F.; Chasteen, L. A.; Chertkov, O.; Chi, H. C.; Christensen, M.; Clark, L. M.; Cohn, J. D.; Denys, M.; Detter, J. C.; Dickson, M.; Dimitrijevic-Bussod, M.; Escobar, J.; Fawcett, J. J.; Flowers, D.; Fotopulos, D.; Glavina, T.; Gomez, M.; Gonzales, E.; Goodstein, D.; Goodwin, L. A.; Grady, D. L.; Grigoriev, I.; Groza, M.; Hammon, N.; Hawkins, T.; Haydu, L.; Hildebrand, C. E.; Huang, W.; Israni, S.; Jett, J.; Jewett, P. B.; Kadner, K.; Kimball, H.; Kobayashi, A.; Krawczyk, M. C.; Leyba, T.; Longmire, J. L.; Lopez, F.; Lou, Y.; Lowry, S.; Ludeman, T.; Manohar, C. F.; Mark, G. A.; McMurray, K. L.; Meincke, L. J.; Morgan, J.; Mozyris, R. K.; Mundt, M. O.; Munk, A. C.; Nandkeshwar, R. D.; Pitluck, S.; Pollard, M.; Predki, P.; Parson-Quintana, B.; Ramirez, L.; Rash, S.; Retterer, J.; Ricke, D. O.; Robinson, D. L.; Rodriguez, A.; Salamov, A.; Saunders, E. H.; Scott, D.; Shough, T.; Stallings, R. L.; Stalvey, M.; Sutherland, R. D.; Tapia, R.; Tesmer, J. G.; Thayer, N.; Thompson, L. S.; Tice, H.; Torney, D. C.; Tran-Gyamfi, M.; Tsai, M.; Ulanovsky, L. E.; Ustaszewska, A.; Vo, N.; White, P. S.; Williams, A. L.; Wills, P. L.; Wu, J. R.; Wu, K.; Yang, J.; DeJong, P.; Bruce, D.; Doggett, N. A.; Deaven, L.; Schmutz, J.; Grimwood, J.; Richardson, P.; Rokhsar, D. S.; Eichler, E. E.; Gilna, P.; Lucas, S. M.; Myers, R. M.; Rubin, E. M.; Pennacchio, L. A. *Nature* **2004**, *432*, 988–994.
- (379) Hillier, L. W.; Graves, T. A.; Fulton, R. S.; Fulton, L. A.; Pepin, K. H.; Minx, P.; Wagner-McPherson, C.; Layman, D.; Wylie, K.; Sekhon, M.; Becker, M. C.; Fewell, G. A.; Delehaunty, K. D.; Miner, T. L.; Nash, W. E.; Kremitzki, C.; Oddy, L.; Du, H.; Sun, H.; Bradshaw-Cordum, H.; Ali, J.; Carter, J.; Cordes, M.; Harris, A.; Isak, A.; van Brunt, A.; Nguyen, C.; Du, F.; Courtney, L.; Kalicki, J.; Ozersky, P.; Abbott, S.; Armstrong, J.; Belter, E. A.; Caruso, L.; Cedroni, M.; Cotton, M.; Davidson, T.; Desai, A.; Elliott, G.; Erb, T.; Fronick, C.; Gaige, T.; Haakenson, W.; Haglund, K.; Holmes, A.; Harkins, R.; Kim, K.; Kruchowski, S. S.; Strong, C. M.; Grewal, N.; Gokye, E.; Hou, S.; Levy, A.; Martinka, S.; Mead, K.; McLellan, M. D.; Meyer, R.; Randall-Maher, J.; Tomlinson, C.; Dauphin-Kohlberg, S.; Kozlowicz-Reilly, A.; Shah, N.; Swarengen-Shahid, S.; Snider, J.; Strong, J. T.; Thompson, J.; Yoakum, M.; Leonard, S.; Pearnan, C.; Trani, L.; Radionenko, M.; Waligorski, J. E.; Wang, C.; Rock, S. M.; Tin-Wollam, A. M.; Maupin, R.; Latreille, P.; Wendl, M. C.; Yang, S. P.; Pohl, C.; Wallis, J. W.; Spieth, J.; Bieri, T. A.; Berkowicz, N.; Nelson, J. O.; Osborne, J.; Ding, L.; Meyer, R.; Sabo, A.; Shotland, Y.; Sinha, P.; Wohldmann, P. E.; Cook, L. L.; Hickenbotham, M. T.; Eldred, J.; Williams, D.; Jones, T. A.; She, X.; Ciccarelli, F. D.; Izaurralde, E.; Taylor, J.; Schmutz, J.; Myers, R. M.; Cox, D. R.; Huang, X.; McPherson, J. D.; Mardis, E. R.; Clifton, S. W.; Warren, W. C.; Chinwalla, A. T.; Eddy, S. R.; Marra, M. A.; Ovcharenko, I.; Furey, T. S.; Miller, W.; Eichler, E. E.; Bork, P.; Suyama, M.; Torrents, D.; Waterston, R. H.; Wilson, R. K. *Nature* **2005**, *434*, 724–731.
- (380) Nusbaum, C.; Zody, M. C.; Borowsky, M. L.; Kamal, M.; Kodira, C. D.; Taylor, T. D.; Whittaker, C. A.; Chang, J. L.; Cuomo, C. A.; Dewar, K.; FitzGerald, M. G.; Yang, X.; Abouelleil, A.; Allen, N. R.; Anderson, S.; Bloom, T.; Bugalter, B.; Butler, J.; Cook, A.; DeCaprio, D.; Engels, R.; Garber, M.; Gnirke, A.; Hafez, N.; Hall, J. L.; Norman, C. H.; Itoh, T.; Jaffe, D. B.; Kuroki, Y.; Lehoczky,

- J.; Lui, A.; Macdonald, P.; Mauceli, E.; Mikkelsen, T. S.; Naylor, J. W.; Nicol, R.; Nguyen, C.; Noguchi, H.; O'Leary, S. B.; O'Neill, K.; Piqani, B.; Smith, C. L.; Talamas, J. A.; Topham, K.; Totoki, Y.; Toyoda, A.; Wain, H. M.; Young, S. K.; Zeng, Q.; Zimmer, A. R.; Fujiyama, A.; Hattori, M.; Birren, B. W.; Sakaki, Y.; Lander, E. S. *Nature* **2005**, *437*, 551–555.
- (381) Nusbaum, C.; Mikkelsen, T. S.; Zody, M. C.; Asakawa, S.; Taudien, S.; Garber, M.; Kodira, C. D.; Schueler, M. G.; Shimizu, A.; Whittaker, C. A.; Chang, J. L.; Cuomo, C. A.; Dewar, K.; FitzGerald, M. G.; Yang, X.; Allen, N. R.; Anderson, S.; Asakawa, T.; Blechschmidt, K.; Bloom, T.; Borowsky, M. L.; Butler, J.; Cook, A.; Corum, B.; DeArellano, K.; DeCaprio, D.; Dooley, K. T.; Dorris, L., III; Engels, R.; Glockner, R. C.; Hafez, N.; Hagopian, D. S.; Hall, J. L.; Ishikawa, S. K.; Jaffe, D. B.; Kamat, A.; Kudoh, J.; Lehmann, R.; Lokitsang, T.; Macdonald, P.; Major, J. E.; Matthews, C. D.; Mauceli, E.; Menzel, U.; Mihalev, A. H.; Minoshima, S.; Murayama, Y.; Naylor, J. W.; Nicol, R.; Nguyen, C.; O'Leary, S. B.; O'Neill, K.; Parker, S. C.; Polley, A.; Raymond, C. K.; Reichwald, K.; Rodriguez, J.; Sasaki, T.; Schilhabel, M.; Siddiqui, R.; Smith, C. L.; Sneddon, T. P.; Talamas, J. A.; Tenzin, P.; Topham, K.; Venkataraman, V.; Wen, G.; Yamazaki, S.; Young, S. K.; Zeng, Q.; Zimmer, A. R.; Rosenthal, A.; Birren, B. W.; Platzer, M.; Shimizu, N.; Lander, E. S. *Nature* **2006**, *439*, 331–335.
- (382) Scherer, S. E.; Muzny, D. M.; Buhay, C. J.; Chen, R.; Cree, A.; Ding, Y.; Dugan-Rocha, S.; Gill, R.; Gunaratne, P.; Harris, R. A.; Hawes, A. C.; Hernandez, J.; Hodgson, A. V.; Hume, J.; Jackson, A.; Khan, Z. M.; Kovar-Smith, C.; Lewis, L. R.; Lozado, R. J.; Metzker, M. L.; Milosavljevic, A.; Miner, G. R.; Montgomery, K. T.; Morgan, M. B.; Nazareth, L. V.; Scott, G.; Sodergren, E.; Song, X. Z.; Steffen, D.; Lovering, R. C.; Wheeler, D. A.; Worley, K. C.; Yuan, Y.; Zhang, Z.; Adams, C. Q.; Ansari-Lari, M. A.; Ayele, M.; Brown, M. J.; Chen, G.; Chen, Z.; Clerc-Blankenburg, K. P.; Davis, C.; Delgado, O.; Dinh, H. H.; Draper, H.; Gonzalez-Garay, M. L.; Havlak, P.; Jackson, L. R.; Jacob, L. S.; Kelly, S. H.; Li, L.; Li, Z.; Liu, J.; Liu, W.; Lu, J.; Maheshwari, M.; Nguyen, B. V.; Okwuonu, G. O.; Pasternak, S.; Perez, L. M.; Plopper, F. J. H.; Santibanez, J.; Shen, H.; Tabor, P. E.; Verduzco, D.; Waldron, L.; Wang, Q.; Williams, G. A.; Zhang, J.; Zhou, J.; Allen, C. C.; Amin, A. G.; Anyalebechi, V.; Bailey, M.; Barbaria, J. A.; Bimage, K. E.; Bryant, N. P.; Burch, P. E.; Burkett, C. E.; Burrell, K. L.; Calderon, E.; Cardenas, V.; Carter, K.; Casias, K.; Cavazos, I.; Cavazos, S. R.; Ceasar, H.; Chacko, J.; Chan, S. N.; Chavez, D.; Christopoulos, C.; Chu, J.; Cockrell, R.; Cox, C. D.; Dang, M.; Dathorne, S. R.; David, R.; Davis, C. M.; Davy-Carroll, L.; Deshazo, D. R.; Donlin, J. E.; D'Souza, L.; Eaves, K. A.; Egan, A.; Emery-Cohen, A. J.; Escotto, M.; Flagg, N.; Forbes, L. D.; Gabisi, A. M.; Garza, M.; Hamilton, C.; Henderson, N.; Hernandez, O.; Hines, S.; Hogues, M. E.; Huang, M.; Idlebird, D. G.; Johnson, R.; Jolivet, A.; Jones, S.; Kagan, R.; King, L. M.; Leal, B.; Lebow, H.; Lee, S.; LeVan, J. M.; Lewis, L. C.; London, P.; Lorensuhewa, L. M.; Loulseged, H.; Lovett, D. A.; Lucier, A.; Lucier, R. L.; Ma, J.; Madu, R. C.; Mapua, P.; Martindale, A. D.; Martinez, E.; Massey, E.; Mawhiney, S.; Meador, M. G.; Mendez, S.; Mercado, C.; Mercado, I. C.; Merritt, C. E.; Miner, Z. L.; Minja, E.; Mitchell, T.; Mohabbat, F.; Mohabbat, K.; Montgomery, B.; Moore, N.; Morris, S.; Munidasa, M.; Ngo, R. N.; Nguyen, N. B.; Nickerson, E.; Nwaokemele, O. O.; Nwokenkwo, S.; Obregon, M.; Oguh, M.; Oragunye, N.; Oviedo, R. J.; Parish, B. J.; Parker, D. N.; Parrish, J.; Parks, K. L.; Paul, H. A.; Payton, B. A.; Perez, A.; Perrin, W.; Pickens, A.; Primus, E. L.; Pu, L. L.; Puazo, M.; Quiles, M. M.; Quiroz, J. B.; Rabata, D.; Reeves, K.; Ruiz, S. J.; Shao, H.; Sisson, I.; Sonaika, T.; Sorelle, R. P.; Sutton, A. E.; Svatek, A. F.; Svetz, L. A.; Tamerisa, K. S.; Taylor, T. R.; Teague, B.; Thomas, N.; Thorn, R. D.; Trejos, Z. Y.; Trevino, B. K.; Ukegbu, O. N.; Urban, J. B.; Vasquez, L. I.; Vera, V. A.; Villasana, D. M.; Wang, L.; Ward-Moore, S.; Warren, J. T.; Wei, X.; White, F.; Williamson, A. L.; Wleczyk, R.; Wooden, H. S.; Wooden, S. H.; Yen, J.; Yoon, L.; Yoon, V.; Zorrilla, S. E.; Nelson, D.; Kucherlapati, R.; Weinstock, G.; Gibbs, R. A. *Nature* **2006**, *440*, 346–351.
- (383) Taylor, T. D.; Noguchi, H.; Totoki, Y.; Toyoda, A.; Kuroki, Y.; Dewar, K.; Lloyd, C.; Itoh, T.; Takeda, T.; Kim, D. W.; She, X.; Barlow, K. F.; Bloom, T.; Bruford, E.; Chang, J. L.; Cuomo, C. A.; Eichler, E.; FitzGerald, M. G.; Jaffe, D. B.; LaButti, K.; Nicol, R.; Park, H. S.; Seaman, C.; Sougnez, C.; Yang, X.; Zimmer, A. R.; Zody, M. C.; Birren, B. W.; Nusbaum, C.; Fujiyama, A.; Hattori, M.; Rogers, J.; Lander, E. S.; Sakaki, Y. *Nature* **2006**, *440*, 497–500.
- (384) Zody, M. C.; Garber, M.; Sharpe, T.; Young, S. K.; Rowen, L.; O'Neill, K.; Whittaker, C. A.; Kamal, M.; Chang, J. L.; Cuomo, C. A.; Dewar, K.; FitzGerald, M. G.; Kodira, C. D.; Madan, A.; Qin, S.; Yang, X.; Abbasi, N.; Abouelleil, A.; Arachchi, H. M.; Baradarani, L.; Birditt, B.; Bloom, S.; Bloom, T.; Borowsky, M. L.; Burke, J.; Butler, J.; Cook, A.; DeArellano, K.; DeCaprio, D.; Dorris, L., III; Dors, M.; Eichler, E. E.; Engels, R.; Fahey, J.; Fleetwood, P.; Friedman, C.; Gearin, G.; Hall, J. L.; Hensley, G.; Johnson, E.; Jones, C.; Kamat, A.; Kaur, A.; Locke, D. P.; Madan, A.; Munson, G.; Jaffe, D. B.; Lui, A.; Macdonald, P.; Mauceli, E.; Naylor, J. W.; Nesbitt, R.; Nicol, R.; O'Leary, S. B.; Ratcliffe, A.; Rounsley, S.; She, X.; Sneddon, K. M. B.; Stewart, S.; Sougnez, C.; Stone, S. M.; Topham, K.; Vincent, D.; Wang, S.; Zimmer, A. R.; Birren, B. W.; Hood, L.; Lander, E. S.; Nusbaum, C. *Nature* **2006**, *440*, 671–675.
- (385) Zody, M. C.; Garber, M.; Adams, D. J.; Sharpe, T.; Harrow, J.; Lupski, J. R.; Nicholson, C.; Searle, S. M.; Wilming, L.; Young, S. K.; Abouelleil, A.; Allen, N. R.; Bi, W.; Bloom, T.; Borowsky, M. L.; Bugalter, B. E.; Butler, J.; Chang, J. L.; Chen, C. K.; Cook, A.; Corum, B.; Cuomo, C. A.; de Jong, P. J.; DeCaprio, D.; Dewar, K.; FitzGerald, M.; Gilbert, J.; Gibson, R.; Gnerre, S.; Goldstein, S.; Graffham, D. V.; Grocock, R.; Hafez, N.; Hagopian, D. S.; Hart, E.; Norman, C. H.; Humphray, S.; Jaffe, D. B.; Jones, M.; Kamal, M.; Khodiyar, V. K.; LaButti, K.; Laird, G.; Lehoczy, J.; Liu, X.; Lokyitsang, T.; Loveland, J.; Lui, A.; Macdonald, P.; Major, J. E.; Matthews, L.; Mauceli, E.; McCarroll, S. A.; Mihalev, A. H.; Mudge, J.; Nguyen, C.; Nicol, R.; O'Leary, S. B.; Osoegawa, K.; Schwartz, D. C.; Shaw-Smith, C.; Stankiewicz, P.; Steward, C.; Swarbreck, D.; Venkataraman, V.; Whittaker, C. A.; Yang, X.; Zimmer, A. R.; Bradley, A.; Hubbard, T.; Birren, B. W.; Rogers, J.; Lander, E. S.; Nusbaum, C. *Nature* **2006**, *440*, 1045–1049.
- (386) Muzny, D. M.; Scherer, S. E.; Kaul, R.; Wang, J.; Yu, J.; Sudbrak, R.; Buhay, C. J.; Chen, R.; Cree, A.; Ding, Y.; Dugan-Rocha, S.; Gill, R.; Gunaratne, P.; Harris, R. A.; Hawes, A. C.; Hernandez, J.; Hodgson, A. V.; Hume, J.; Jackson, A.; Khan, Z. M.; Kovar-Smith, C.; Lewis, L. R.; Lozado, R. J.; Metzker, M. L.; Milosavljevic, A.; Miner, G. R.; Morgan, M. B.; Nazareth, L. V.; Scott, G.; Sodergren, E.; Song, X. Z.; Steffen, D.; Wei, S.; Wheeler, D. A.; Wright, M. W.; Worley, K. C.; Yuan, Y.; Zhang, Z.; Adams, C. Q.; Ansari-Lari, M. A.; Ayele, M.; Brown, M. J.; Chen, G.; Chen, Z.; Clendenning, J.; Clerc-Blankenburg, K. P.; Chen, R.; Chen, Z.; Davis, C.; Delgado, O.; Dinh, H. H.; Dong, W.; Draper, H.; Ernst, S.; Fu, G.; Gonzalez-Garay, M. L.; Garcia, D. K.; Gillett, W.; Gu, J.; Hao, B.; Haugen, E.; Havlak, P.; He, X.; Hennig, S.; Hu, S.; Huang, W.; Jackson, L. R.; Jacob, L. S.; Kelly, S. H.; Kube, M.; Levy, R.; Li, Z.; Liu, B.; Liu, J.; Liu, W.; Lu, J.; Maheshwari, M.; Nguyen, B. V.; Okwuonu, G. O.; Palmeiri, A.; Pasternak, S.; Perez, L. M.; Phelps, K. A.; Plopper, F. J. H.; Qiang, B.; Raymond, C.; Rodriguez, R.; Saenphimmachak, C.; Santibanez, J.; Shen, H.; Shen, Y.; Subramanian, S.; Tabor, P. E.; Verduzco, D.; Waldron, L.; Wang, J.; Wang, J.; Wang, Q.; Williams, G. A.; Wong, G. K. S.; Yao, Z.; Zhang, J.; Zhang, X.; Zhao, G.; Zhou, J.; Zhou, Y.; Nelson, D.; Lehrach, H.; Reinhardt, R.; Naylor, S. L.; Yang, H.; Olson, M.; Weinstock, G.; Gibbs, R. A. *Nature* **2006**, *440*, 1194–1198.

CR068304C