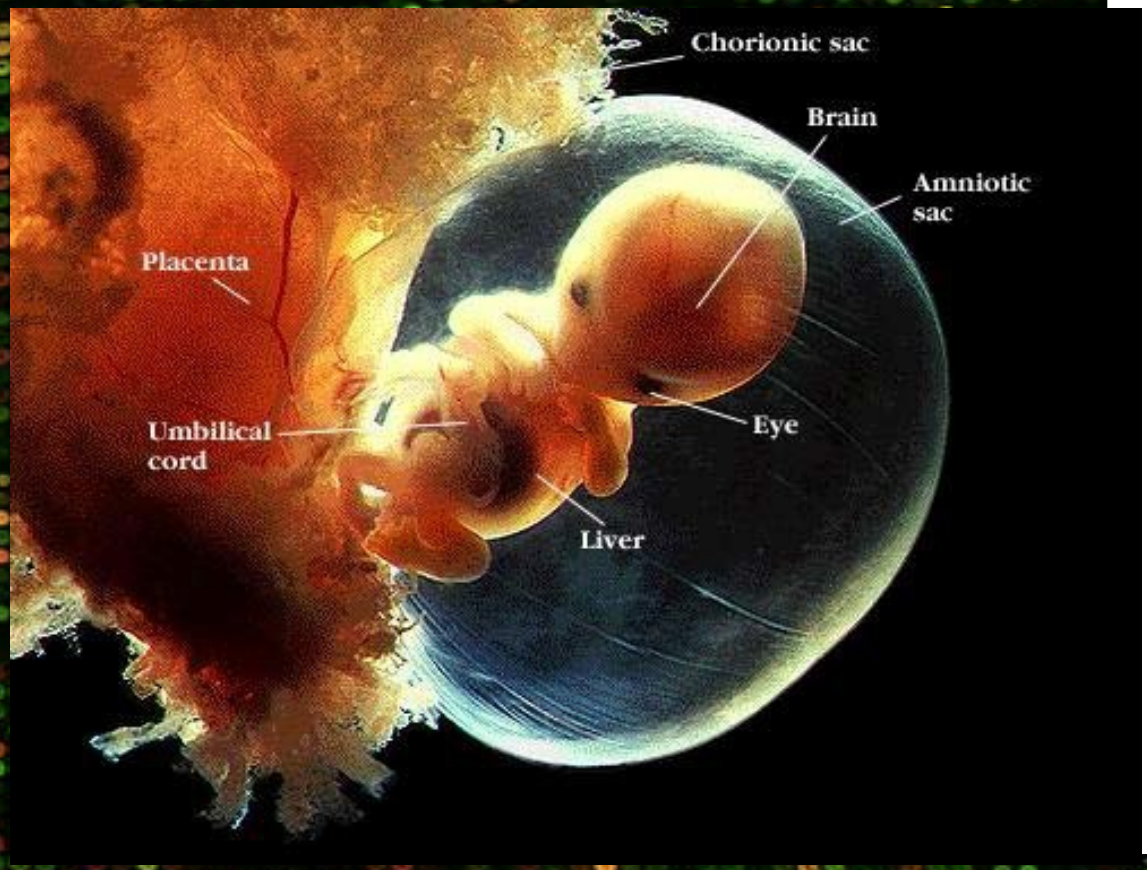
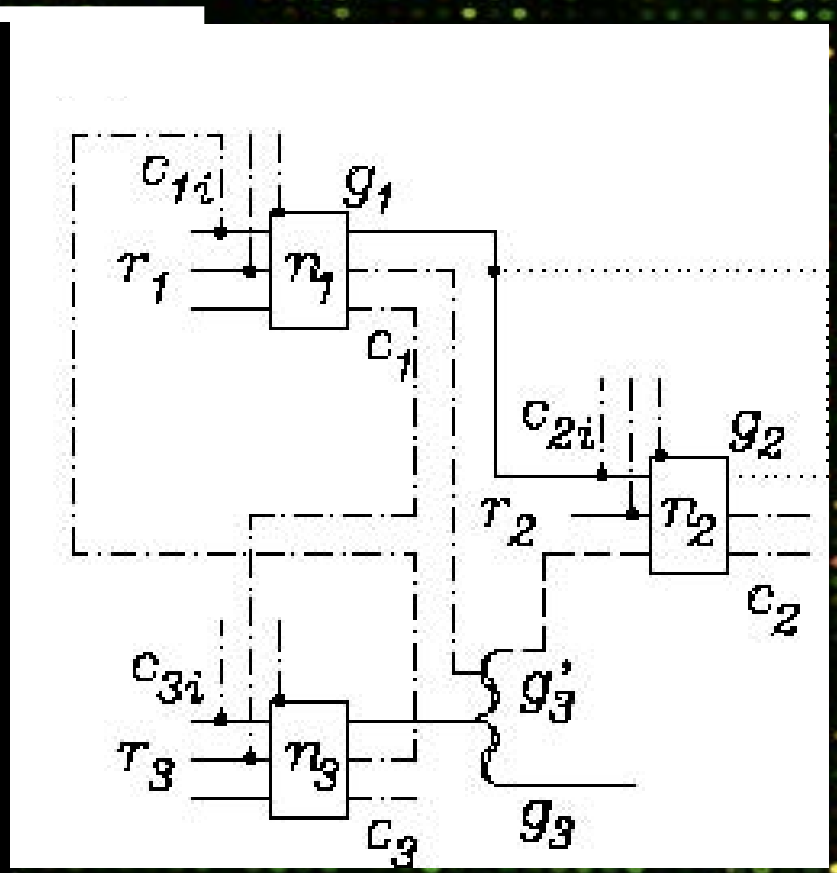
RNA at the epicentre of human evolution and development





John Mattick

Garvan Institute of Medical Research, Sydney

DNA habitats and its RNA inhabitants, Salzburg - 4 July 2014

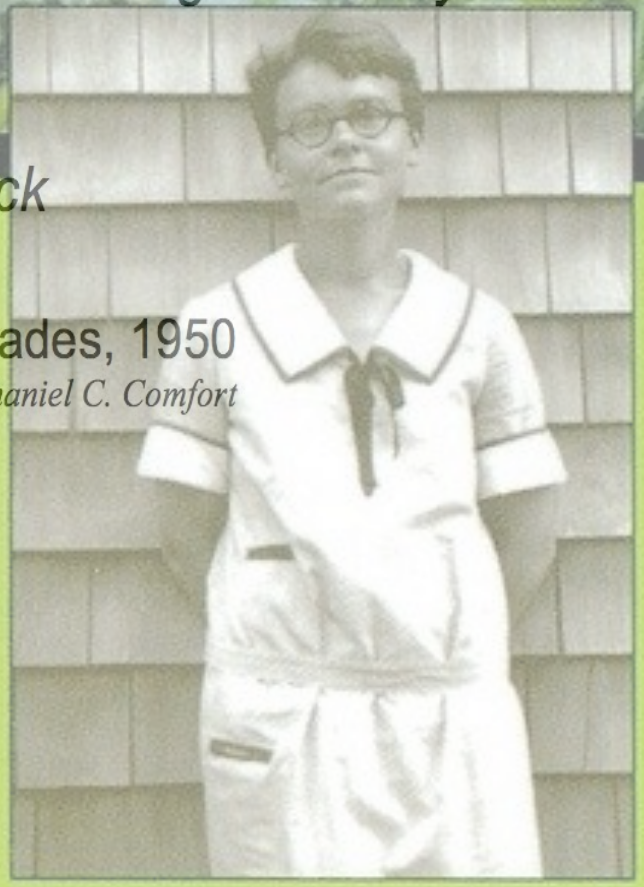
Are we letting a philosophy of the [protein-coding] gene control [our] reasoning? What, then, is the philosophy of the gene? Is it a valid philosophy?

When one starts to question the reasoning behind the origin of the present notion of the gene (held by most geneticists) the opportunity for questioning its validity becomes apparent.

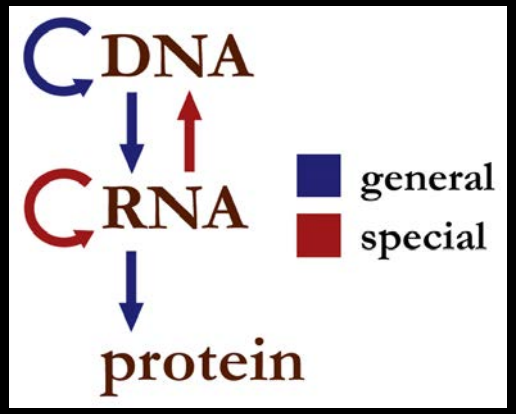
-Barbara McClintock

Letter to Marcus Rhoades, 1950

From The Tangled Field by Nathaniel C. Comfort



The assumption, based on studies of the *lac* operon in *E. coli*, has been that genes are synonymous with proteins and that most genetic information, including regulatory information, is transacted by proteins.



This protein-centric view reflects the mechanical and reductionist zeitgeist of the age and led to several subsidiary assumptions, despite a number of subsequent surprises that should have given pause for thought.

Surprise #1: Genes in humans and other complex eukaryotes are mosaics of protein-coding and noncoding sequences.

Interpretation: Intervening sequences, despite the fact that they are transcribed, are ‘junk’.

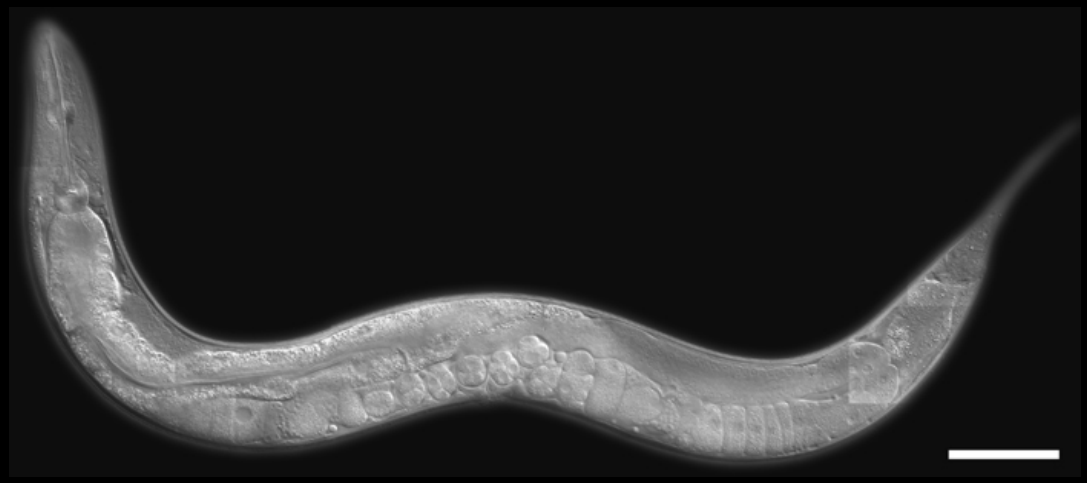
Surprise #2: Eukaryote genomes are full of transposon-derived sequences.

Interpretation: These sequences are mainly non-functional ‘selfish’ DNA (!).

Surprise #3: Gene number does not scale with developmental complexity.

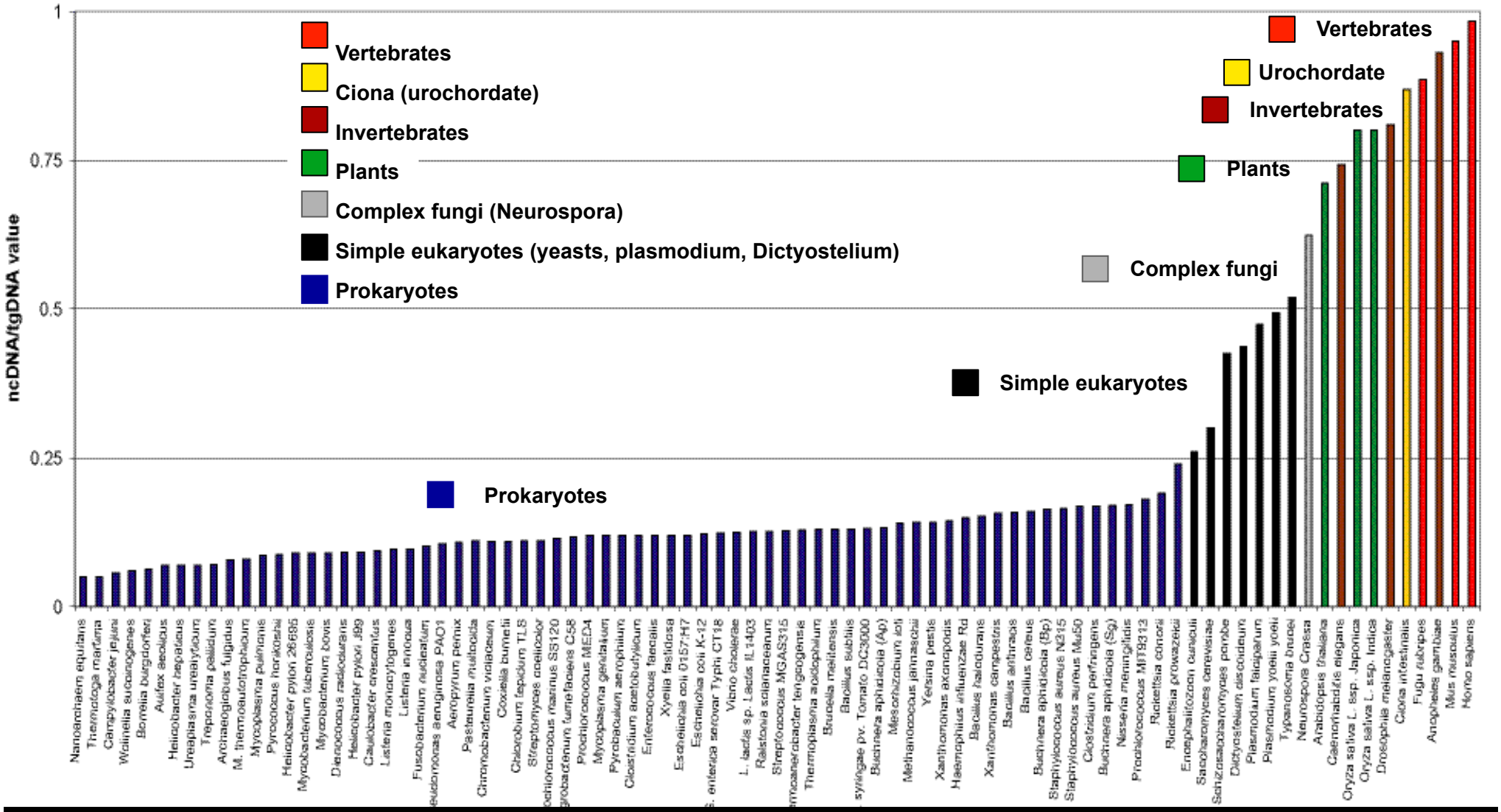
Interpretation: Combinatorial control of transcription, splicing etc. can explain ….?

The genetic basis of human development



- Humans (and other vertebrates) have approximately the same number of protein-coding genes (~20,000) as *C. elegans.*
- Most of the proteins are orthologous and have similar functions in all animals, and many are common with yeast.
- Where is the information that programs our developmental and cognitive complexity?

The proportion of noncoding DNA broadly increases with developmental complexity



Irrespective of the extent of non-coding sequences, it is now evident that the vast majority of the genomes of all organisms is transcribed in a dynamic manner in different cells and tissues at different developmental stages.

The Transcriptional Landscape of the Mammalian Genome

The FANTOM Consortium* and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group)*

This study describes comprehensive polling of transcription start and termination sites and analysis of previously unidentified full-length complementary DNAs derived from the mouse genome. We identify the 5′ and 3′ boundaries of 181,047 transcripts with extensive variation in transcripts arising from alternative promoter usage, splicing, and polyadenylation. There are 16,247 new mouse protein-coding transcripts, including 5154 encoding previously unidentified proteins. Genomic mapping of the transcriptome reveals transcriptional forests, with overlapping transcription on both strands, separated by deserts in which few transcripts are observed. The data provide a comprehensive platform for the comparative analysis of mammalian transcriptional regulation in differentiation and development.

Identified > 30,000 transcripts with little or no protein-coding potential

Antisense Transcription in the Mammalian Transcriptome

RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) and the FANTOM Consortium

Antisense transcription (transcription from the opposite strand to a protein-coding or sense strand) has been ascribed roles in gene regulation involving degradation of the corresponding sense transcripts (RNA interference), as well as gene silencing at the chromatin level. Global transcriptome analysis provides evidence that a large proportion of the genome can produce transcripts from both strands, and that antisense transcripts commonly link neighboring "genes" in complex loci into chains of linked transcriptional units. Expression profiling reveals frequent concordant regulation of sense/antisense pairs. We present experimental evidence that perturbation of an antisense RNA can alter the expression of sense messenger RNAs, suggesting that antisense transcription contributes to control of transcriptional outputs in mammals.

~70% of mouse genes exhibit overlapping antisense transcripts

Transcriptional Maps of 10 Human Chromosomes at 5-Nucleotide Resolution

Jill Cheng,[1*] Philipp Kapranov,[1*] Jorg Drenkow,[1] Sujit Dike,[1] Shane Brubaker,[1] Sandeep Patel,[1] Jeffrey Long,[1] David Stern,[1] Hari Tammana,[1] Gregg Helt,[1] Victor Sementchenko,[1] Antonio Piccolboni,[1] Stefan Bekiranov,[1] Dione K. Bailey,[1] Madhavan Ganesh,[1] Srinka Ghosh,[1] Ian Bell,[1] Daniela S. Gerhard,[2] Thomas R. Gingeras[1]†

Sites of transcription of polyadenylated and nonpolyadenylated RNAs for 10 human chromosomes were mapped at 5–base pair resolution in eight cell lines. Unannotated, nonpolyadenylated transcripts comprise the major proportion of the transcriptional output of the human genome. Of all transcribed sequences, 19.4, 43.7 and 36.9% were observed to be polyadenylated, nonpolyadenylated, and bimorphic, respectively. Half of all transcribed sequences are found only in the nucleus and for the most part are unannotated. Overall, the transcribed portions of the human genome are predominantly composed of interlaced networks of both poly A+ and poly A– annotated transcripts and unannotated transcripts of unknown function. This organization has important implications for interpreting genotype-phenotype associations, regulation of gene expression, and the definition of a gene.

~44% of human transcripts are not polyadenylated and comprise a largely distinct set of sequences

ARTICLE

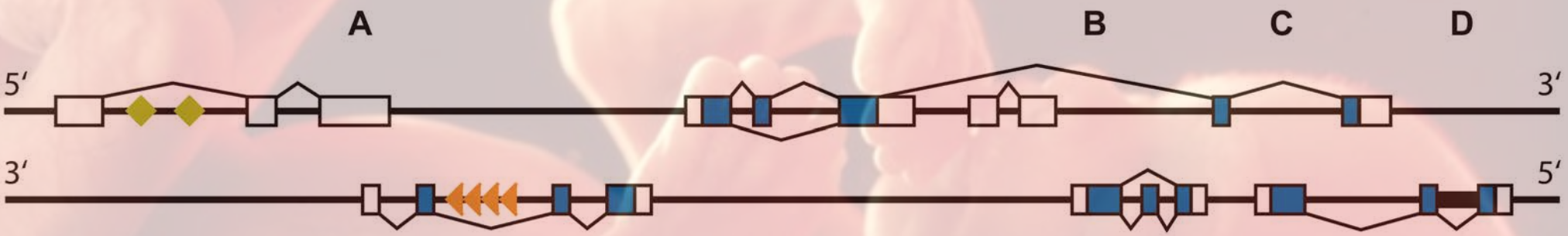
doi:10.1038/nature12535

Genomic organization of human transcription initiation complexes

Bryan J. Venters[1]† & B. Franklin Pugh[1]

The human genome is pervasively transcribed, yet only a small fraction is coding. Here we address whether this non-coding transcription arises at promoters, and detail the interactions of initiation factors TATA box binding protein (TBP), transcription factor IIB (TFIIB) and RNA polymerase (Pol) II. Using ChIP-exo (chromatin immunoprecipitation with lambda exonuclease digestion followed by high-throughput sequencing), we identify approximately 160,000 transcription initiation complexes across the human K562 genome, and more in other cancer genomes. Only about 5% associate with messenger RNA genes. The remainder associates with non-polyadenylated non-coding transcription. Regardless, Pol II moves into a transcriptionally paused state, and TBP and TFIIB remain at the promoter. Remarkably, the vast majority of locations contain the four core promoter elements—upstream TFIIB recognition element (BRE_u), TATA, downstream TFIIB recognition element (BRE_d), and initiator element (INR)—in constrained positions. All but the INR also reside at Pol III promoters, where TBP makes similar contacts. This comprehensive and high-resolution genome-wide detection of the initiation machinery produces a consolidated view of transcription initiation events from yeast to humans at Pol II/III TATA-containing/TATA-less coding and non-coding genes.

The amazing complexity of the mammalian transcriptome



Graphical representation of the complexity of the transcriptional landscape in mammals. **White boxes** represent non-coding exonic sequences and **blue** boxes protein-coding exonic sequences. **Green diamonds** represent snoRNAs and **orange triangles** represent miRNAs. Indicated are (A) antisense transcripts with overlapping exons, (B) nested transcripts on both strands, (C) antisense transcripts with interlacing exons, and (D) retained introns. j

JS Mattick and IV Makunin (2006) Non-coding RNA. *Human Molecular Genetics* 15: R17-R29

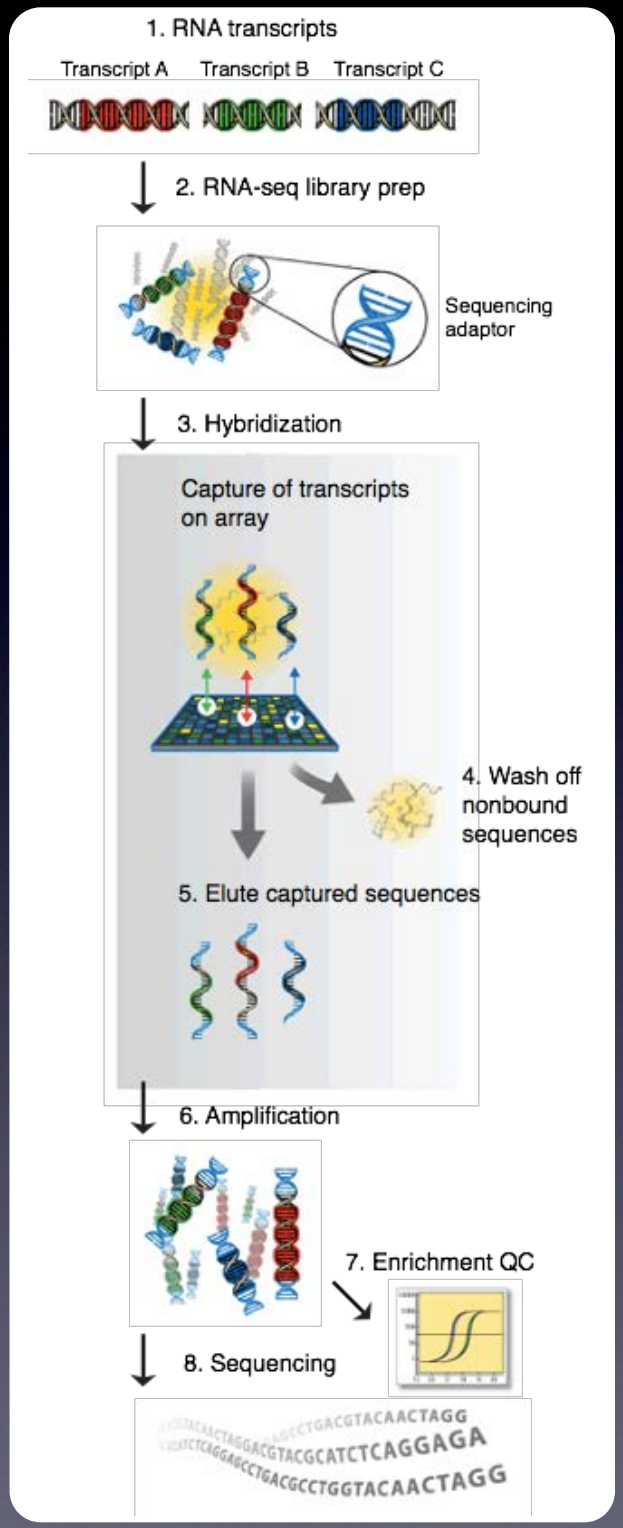
RNA capture-sequencing: focussed transcriptomics a là exome sequencing

- Capture arrays contain probes that hybridize to RNAs expressed from genomic regions of interest

- Transcripts of interest are captured by incubating a RNA sequencing library with the array. Non-target RNAs are washed away.

- Captured transcripts are eluted and enrichment of targeted RNAs confirmed

- Captured transcripts are sequenced

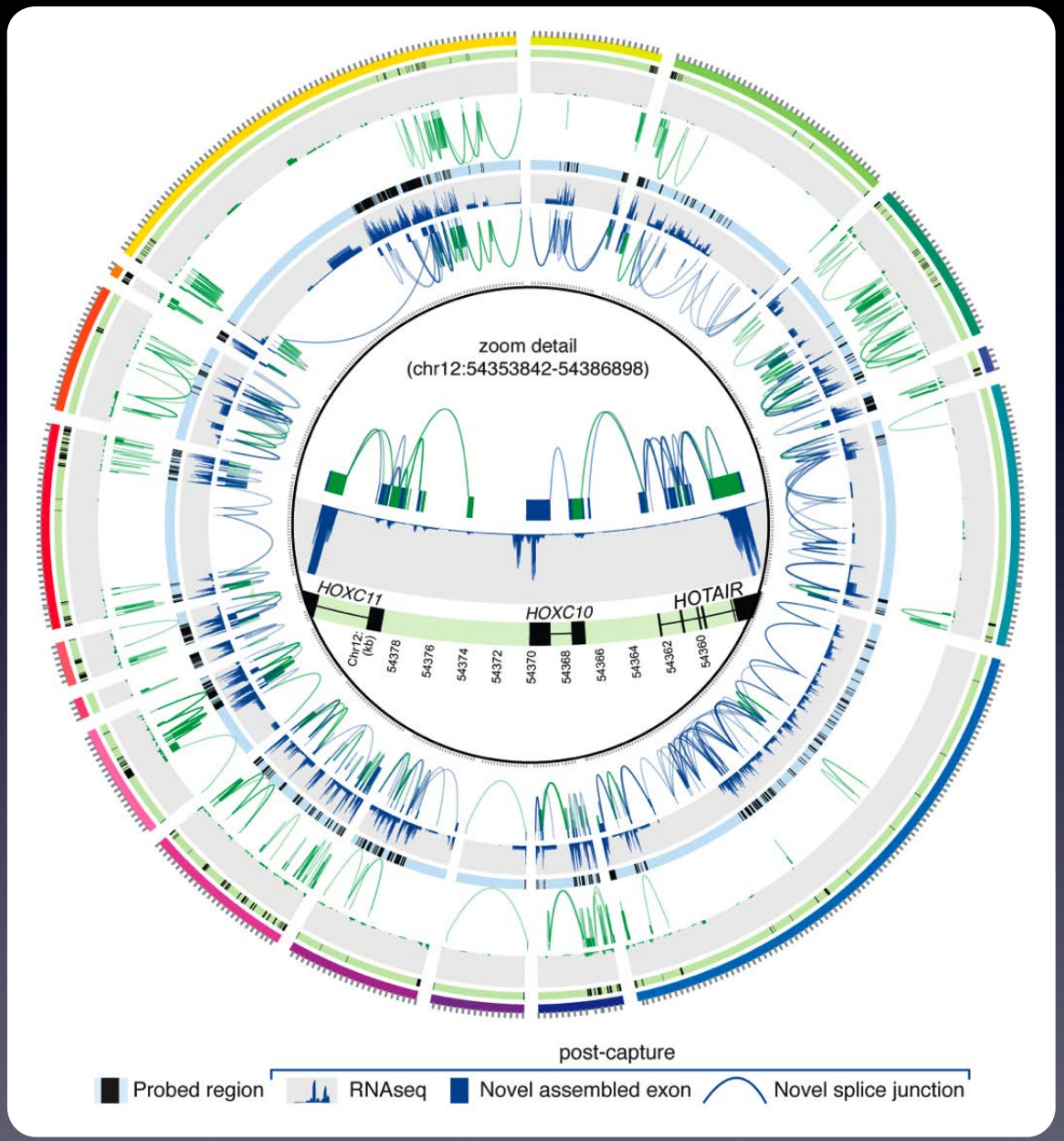


RNA Capture-Seq exposes the deep complexity of the human transcriptome

We used RNA Capture-Seq to examine transcription in intergenic loci that are "gene deserts" as identified by conventional RNA-Seq

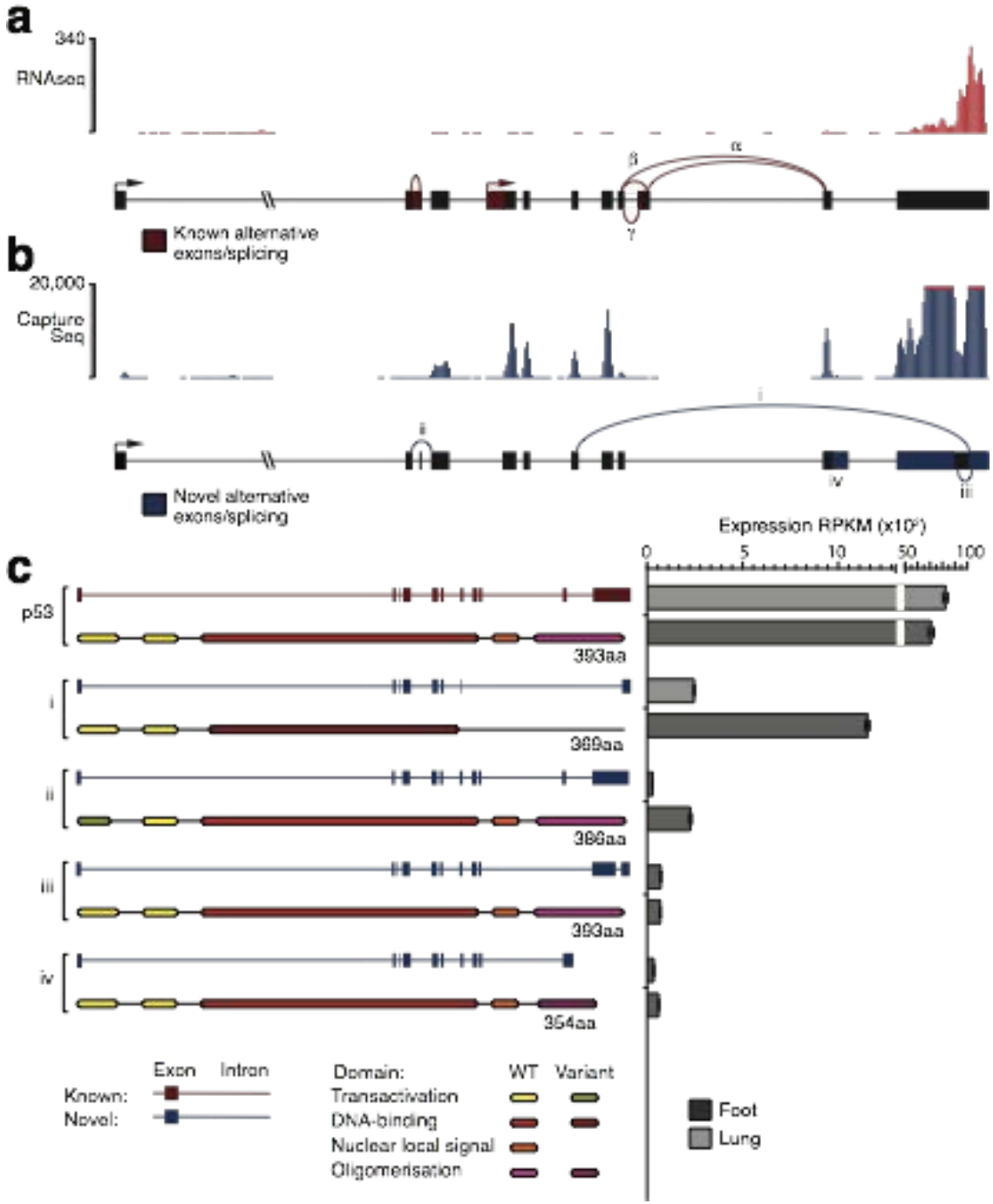
Capture-Seq tranforms regions of sparse-mapping RNA-seq reads into long, complex alternatively spliced RNAs (95% of transcripts in intergenic regions completely novel)

Protein-coding loci are similarly transformed, revealing many previously undetected spliced isoforms

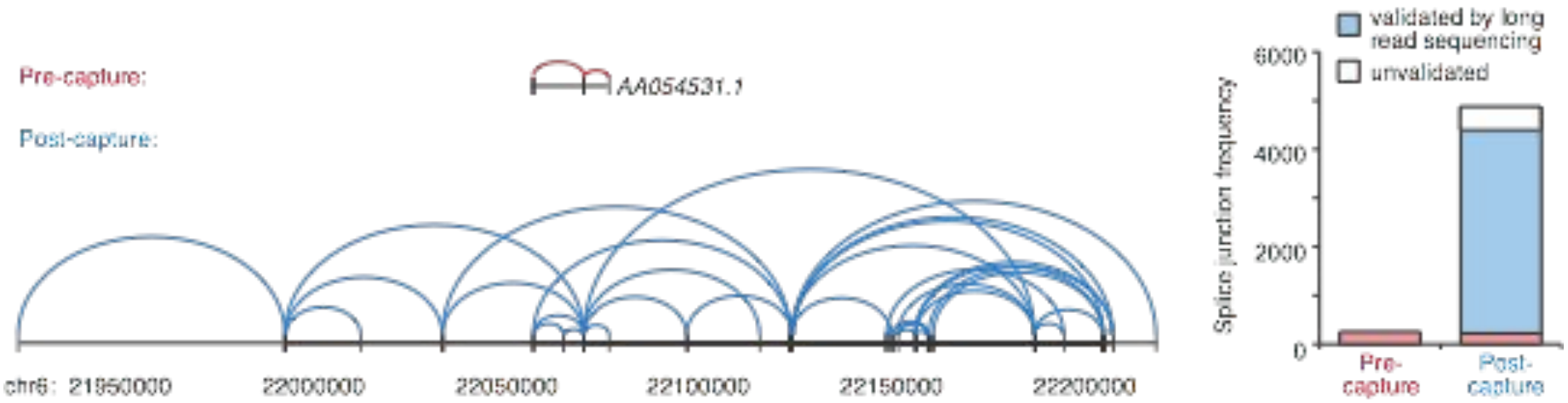


Tim Mercer, Daniel Gerhardt, Marcel Dinger, Cole Trapnell, Jeff Jeddeloh, John Mattick and John Rinn
Nature Biotechnology (2012) **30**: 99-104.

Discovery of 4 new isoforms of p53 by RNA CaptureSeq



Transcriptional and splicing complexity of a noncoding RNA locus revealed by RNA CaptureSeq

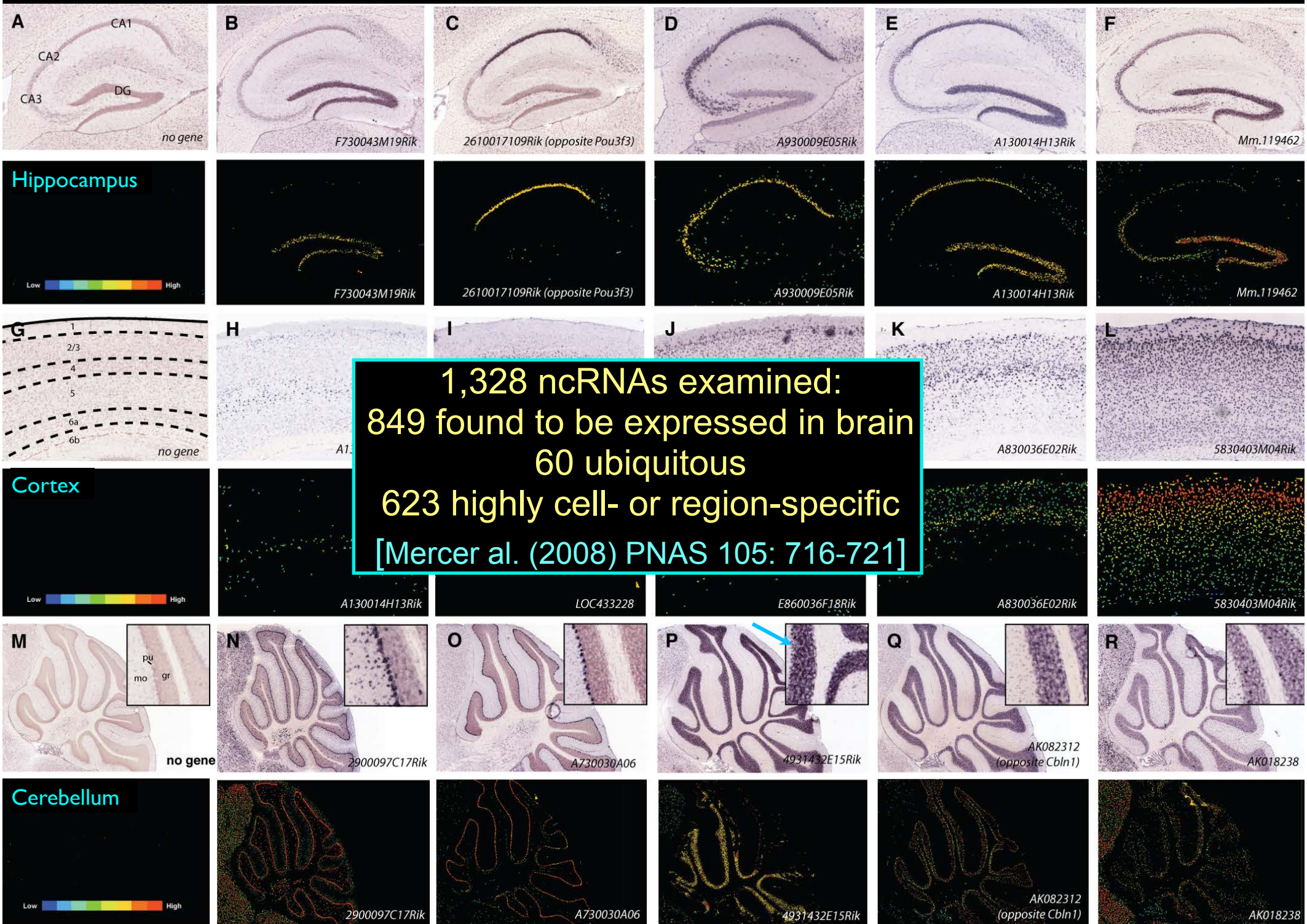


The vast majority of catalogued noncoding RNAs show no evidence of translation

Feature	References
Conservation of promoters	[2,27,32]
Conservation of splice junctions	[27]
Conservation of sequence	[26,27,32]
Conservation of genomic position	[31,33,34]
Conservation of secondary structure	[28–30]
Positive selection	[230]
Conservation of expression	[35,36]
Dynamic expression and alternative splicing	[13,31,32]
Altered expression or splicing in cancer and other diseases	[37–49]
Association with particular chromatin signatures	[31,32]
Regulation by morphogens and transcription factors	[31,32,49,50]
Tissue- and cell-specific expression patterns	[16,17,19–22,49,51–56]
Specific subcellular localization	[19–22,52,56]

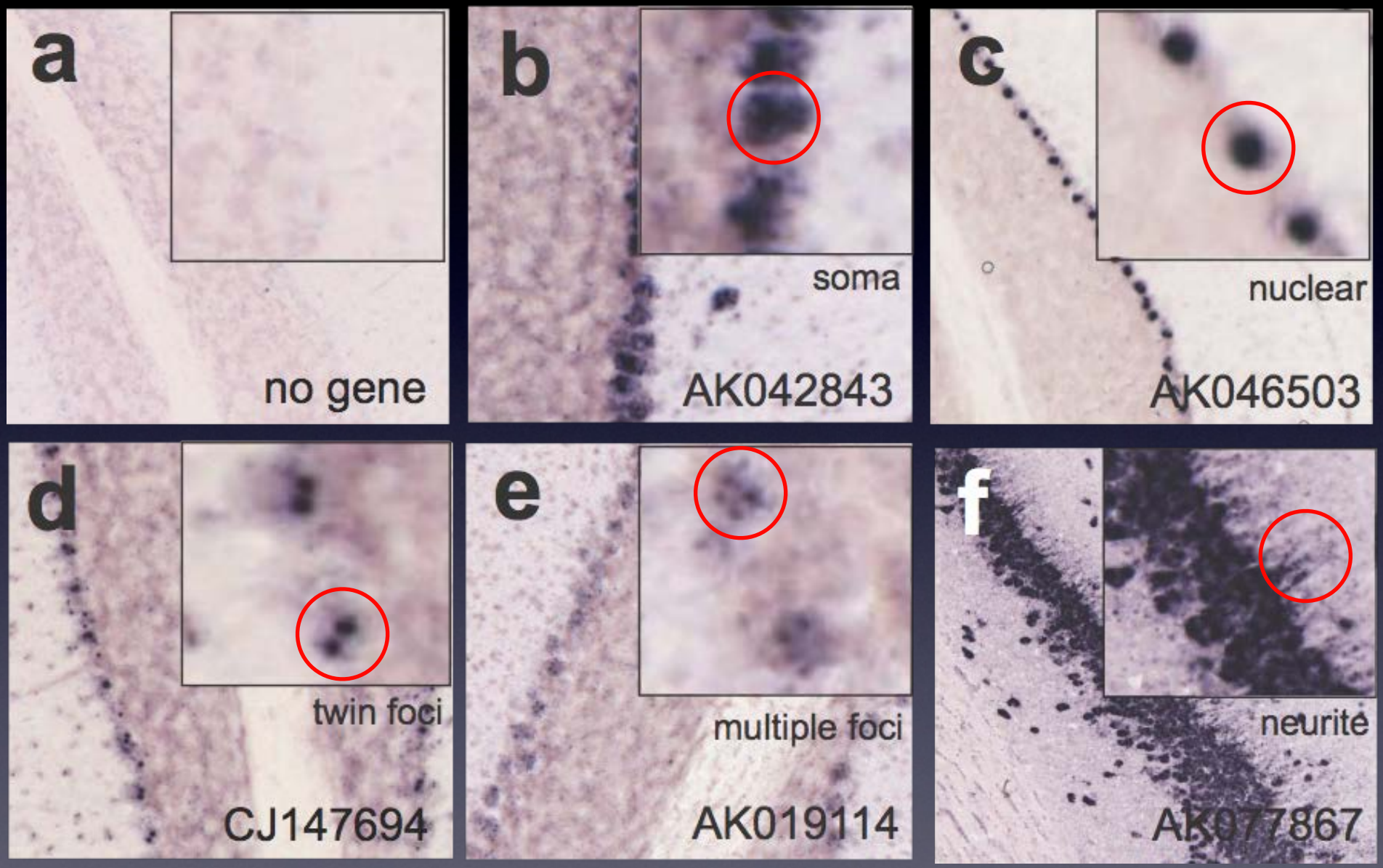
Mattick JS (2009) The genetic signatures of noncoding RNAs. *PLoS Genetics*, e1000459

Non-coding RNA expression in mouse brain



1,328 ncRNAs examined:
849 found to be expressed in brain
60 ubiquitous
623 highly cell- or region-specific
[Mercer al. (2008) PNAS 105: 716-721]

Subcellular localization of long ncRNAs



Mercer, Dinger et al., PNAS 2008

CSH PRESS GENOME RESEARCH

LightCycler® 480 Real-Time PCR System
Get what you need, and more
for real-time PCR applications!
Roche

Also called *Neat1*

MEN ε/β nuclear retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles

Hongjae Sunwoo[1], Marcel E. Dinger[2], Jeremy E. Wilusz[3], Paulo P. Amaral[2], John S. Mattick[2], and David L. Spector[1,4]

Text

+ Author Affiliations

ACCEPTED PREPRINT

This Article

Published in Advance December 22, 2008, doi: 10.1101/gr.087775.108

Genome Res. 2008.

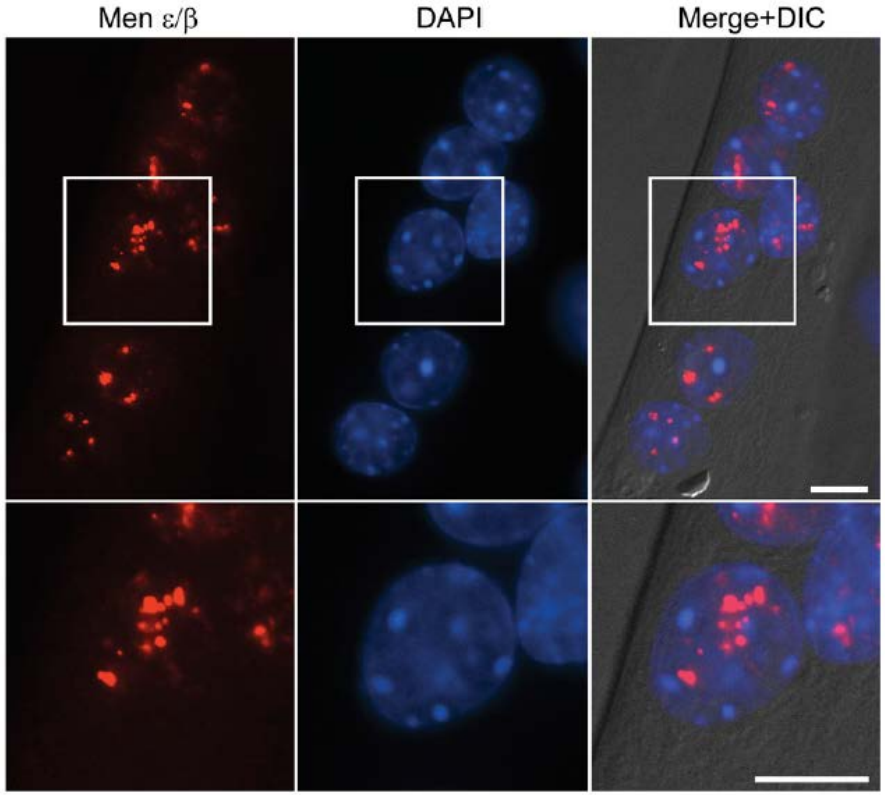
Copyright © 2008, Cold Spring Harbor Laboratory Press

» Abstract *Free*

Current Issue
January 2009, 19 (1)

Abstract

Studies of the transcriptional output of the human and mouse genomes have revealed that there are many more transcripts produced than can be accounted for by predicted protein-coding genes. Using a custom microarray, we have identified 184 non-coding RNAs that exhibit more than 2 fold up- or down-regulation upon differentiation of C2C12 myoblasts into myotubes. Here, we focus on the *Men ε/β* locus, which is up-regulated 3.3 fold during differentiation. Two non-coding RNA isoforms are produced from a single RNA polymerase II promoter, differing in the location of their 3' ends. *Men ε* is a 3.2-kb polyadenylated RNA, whereas *Men β* is a ~20-kb transcript containing a genomically encoded poly(A)-rich tract at its 3' end. The 3' end of *Men β* is generated by RNase P cleavage. The *Men ε/β* transcripts are localized to nuclear paraspeckles and directly interact with NONO. Knock-down of *MEN ε/β* expression results in the disruption of nuclear paraspeckles. Furthermore, the formation of paraspeckles, after release from transcriptional inhibition by DRB treatment, was suppressed in *MEN ε/β* depleted cells. Our findings indicate that the *MEN ε/β* non-coding RNAs are essential structural/organizational components of paraspeckles.

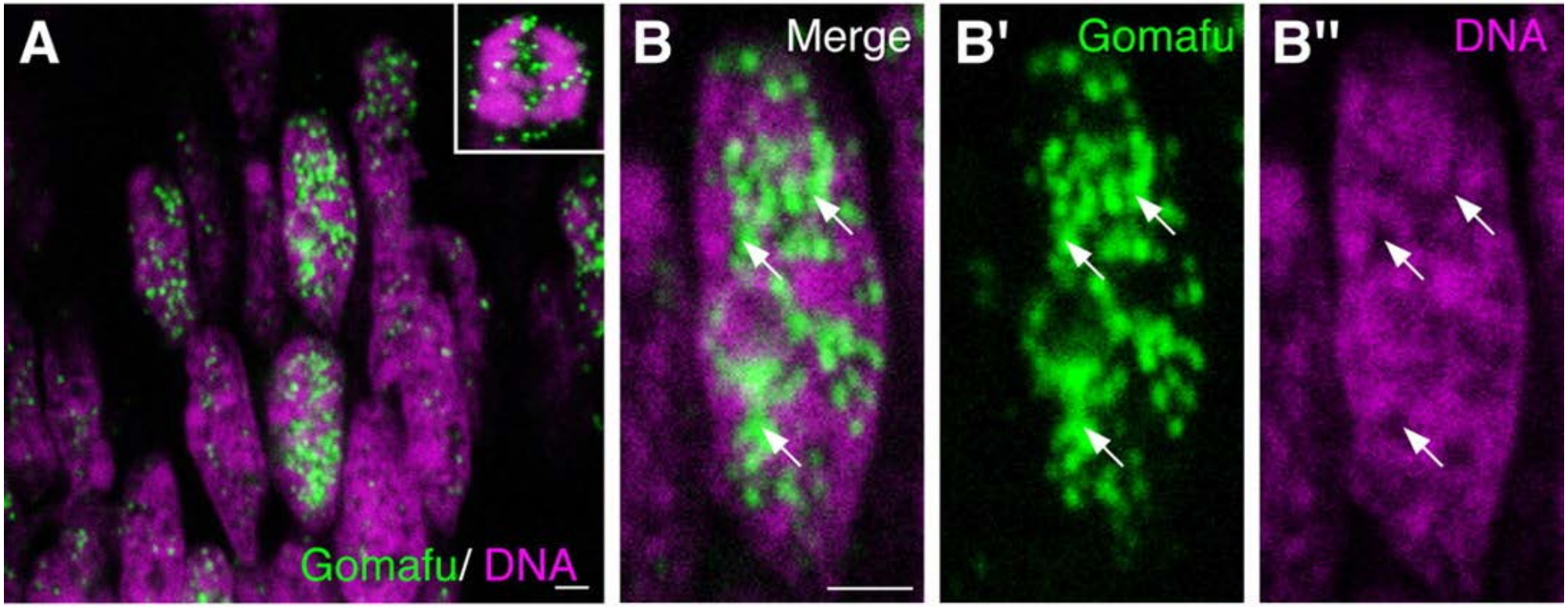


The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons

Masamitsu Sone[1,2,3], Tetsutaro Hayashi[2], Hiroshi Tarui[2], Kiyokazu Agata[2], Masatoshi Takeichi[2,3] and Shinichi Nakagawa[1,2,*]

[1]Nakagawa Initiative Research Unit, RIKEN, 2-1 Hirosawa, Wako 351-0198, Japan
[2]RIKEN Center for Developmental Biology, 2-2-3 Minatojima Minamimachi, Chuo-ku, Kobe 650-0047, Japan
[3]Department of Cell and Developmental Biology, Graduate School of Biostudies, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan
*Author for correspondence (e-mail: nakagawas@riken.jp)

Accepted 21 May 2007
Journal of Cell Science 120, 2498-2506 Published by The Company of Biologists 2007
doi:10.1242/jcs.009957



www.nature.com/mp

ORIGINAL ARTICLE

The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing

G Barry[1], JA Briggs[2], DP Vanichkina[1], EM Poth[3], NJ Beveridge[4,5], VS Ratnu[6], SP Nayler[2], K Nones[7], J Hu[8], TW Bredy[6], S Nakagawa[9], F Rigo[10], RJ Taft[1], MJ Cairns[4,5], S Blackshaw[3], EJ Wolvetang[2] and JS Mattick[1,11,12]

Schizophrenia (SZ) is a complex disease characterized by impaired neuronal functioning. Although defective alternative splicing has been linked to SZ, the molecular mechanisms responsible are unknown. Additionally, there is limited understanding of the early transcriptomic responses to neuronal activation. Here, we profile these transcriptomic responses and show that long non-coding RNAs (lncRNAs) are dynamically regulated by neuronal activation, including acute downregulation of the lncRNA Gomafu, previously implicated in brain and retinal development. Moreover, we demonstrate that Gomafu binds directly to the splicing factors QKI and SRSF1 (serine/arginine-rich splicing factor 1) and dysregulation of Gomafu leads to alternative splicing patterns that resemble those observed in SZ for the archetypal SZ-associated genes *DISC1* and *ERBB4*. Finally, we show that Gomafu is downregulated in post-mortem cortical gray matter from the superior temporal gyrus in SZ. These results functionally link activity-regulated lncRNAs and alternative splicing in neuronal function and suggest that their dysregulation may contribute to neurological disorders.

Molecular Psychiatry advance online publication, 30 April 2013; doi:10.1038/mp.2013.45

Keywords: alternative splicing; Gomafu; neuronal activation; quaking homolog; schizophrenia

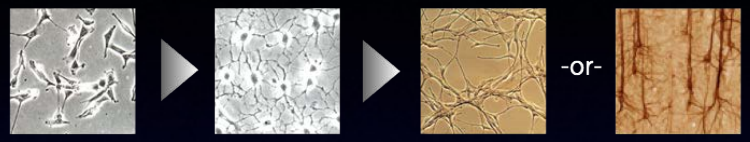
Distinct suites of lncRNAs are expressed at different stages of development

Embryonic stem cell differentiation

Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation

Marcel E. Dinger,[1,6] Paulo P. Amaral,[1,6] Tim R. Mercer,[1,6] Ken C. Pang,[1,2] Stephen J. Bruce,[1] Brooke B. Gardiner,[1,3] Marjan E. Askarian-Amiri,[1] Kelin Ru,[1] Giulia Soldà,[1,4] Cas Simons,[1] Susan M. Sunkin,[5] Mark L. Crowe,[1] Sean M. Grimmond,[1,3] Andrew C. Perkins,[1] and John S. Mattick[1,7]

Genome Research 2008

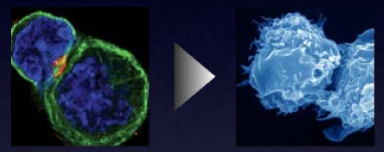


Neuronal stem cell differentiation

Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation

Tim R Mercer[1†], Irfan A Qureshi[2,3,6†], Solen Gokhan[2,3,6], Marcel E Dinger[1], Guangyu Li[3,6], John S Mattick[1*], Mark F Mehler[2,3,4,5,6*]

BMC Neuroscience 2009

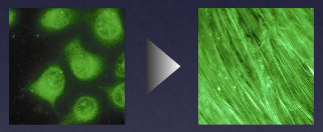


T-cell activation

Genome-Wide Identification of Long Noncoding RNAs in CD8$^+$ T Cells[1]

Ken C. Pang,[2*†] Marcel E. Dinger,* Tim R. Mercer,* Lorenzo Malquori,* Sean M. Grimmond,*‡ Weisan Chen,† and John S. Mattick[3*]

Journal of Immunology 2009

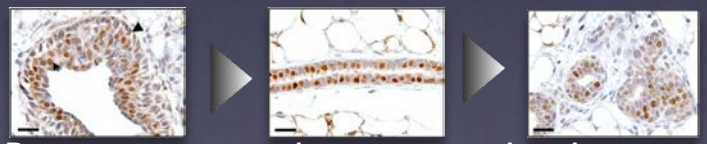


Myoblast differentiation

MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles

Hongjae Sunwoo,[1] Marcel E. Dinger,[2] Jeremy E. Wilusz,[3] Paulo P. Amaral,[2] John S. Mattick,[2] and David L. Spector[1,3,4]

Genome Research 2009

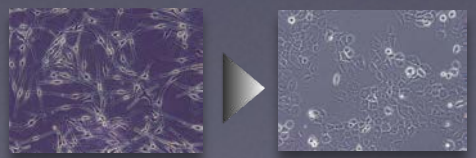


Breast cancer and mammary development

SNORD-host RNA *Zfas1* is a regulator of mammary development and a potential marker for breast cancer

MARJAN E. ASKARIAN-AMIRI,[1] JOANNA CRAWFORD,[1] JULIET D. FRENCH,[2] CHANEL E. SMART,[2,4] MARTIN A. SMITH,[1] MICHAEL B. CLARK,[1] KELIN RU,[1] TIM R. MERCER,[1] ELLA R. THOMPSON,[3] SUNIL R. LAKHANI,[4,5,6] ANA C. VARGAS,[4] IAN G. CAMPBELL,[3,7] MELISSA A. BROWN,[2] MARCEL E. DINGER,[1,8] and JOHN S. MATTICK[1,8]

RNA 2011



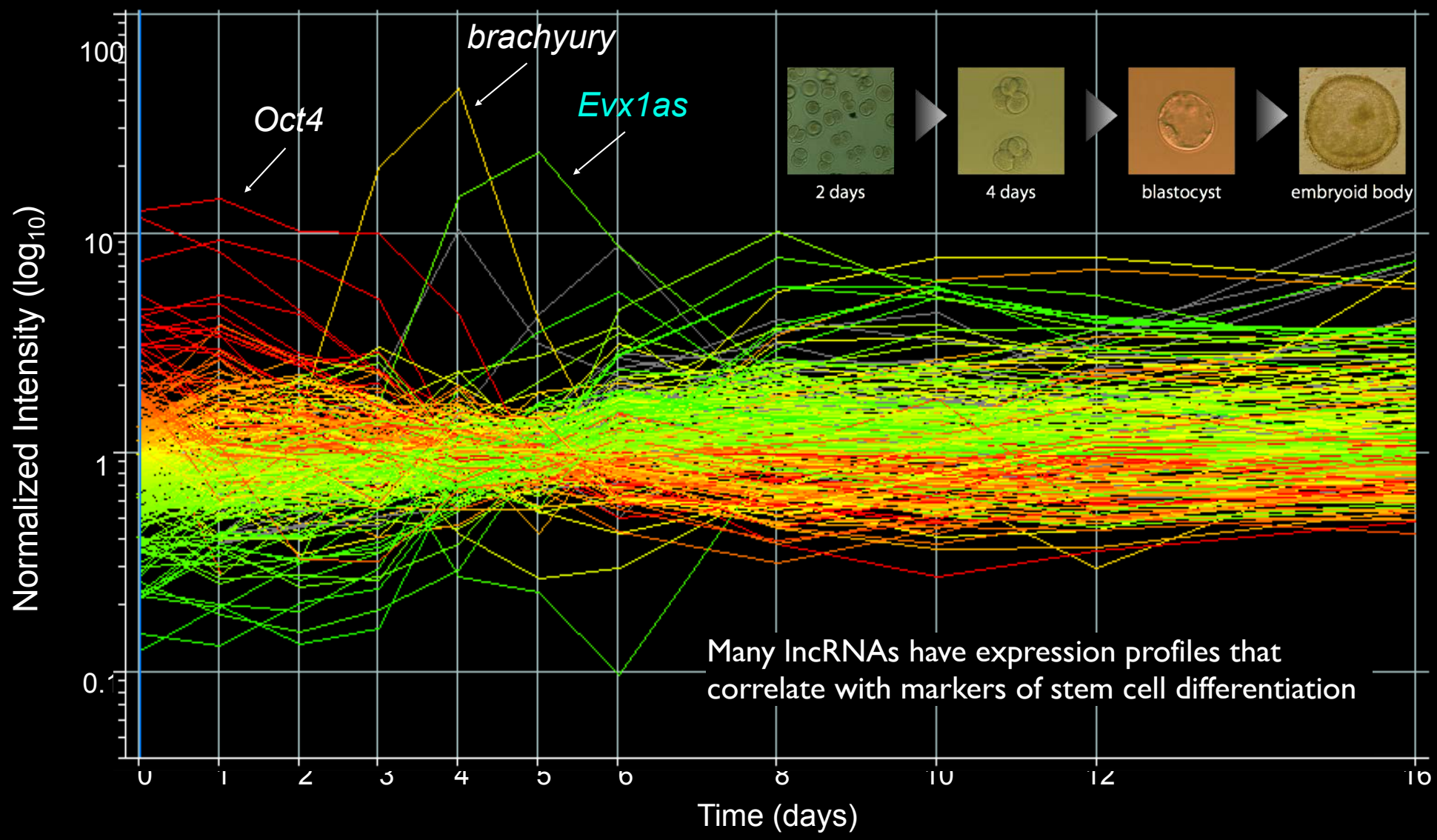
Melanoma

The Melanoma-Upregulated Long Noncoding RNA *SPRY4-IT1* Modulates Apoptosis and Invasion

Divya Khaitan[1], Marcel E. Dinger[2], Joseph Mazar[1], Joanna Crawford[2], Martin A. Smith[2], John S. Mattick[2], and Ranjan J. Perera[1]

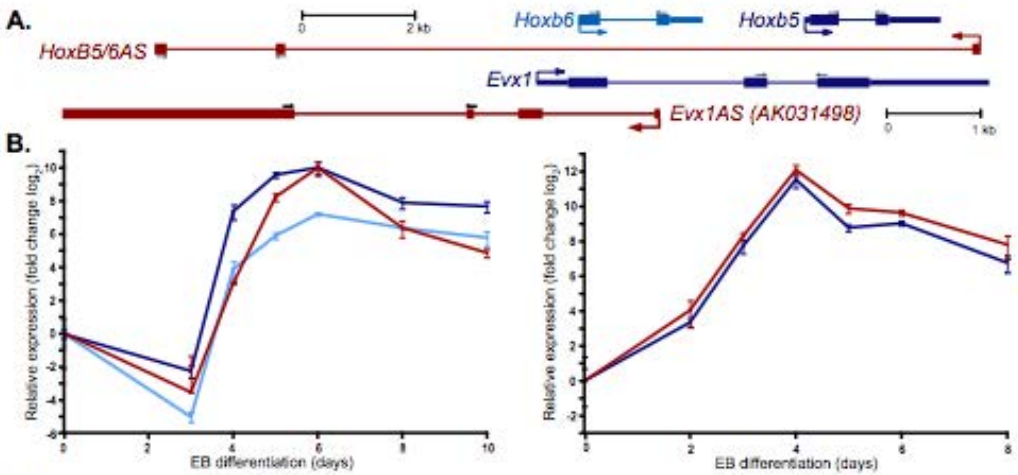
Cancer Research 2011

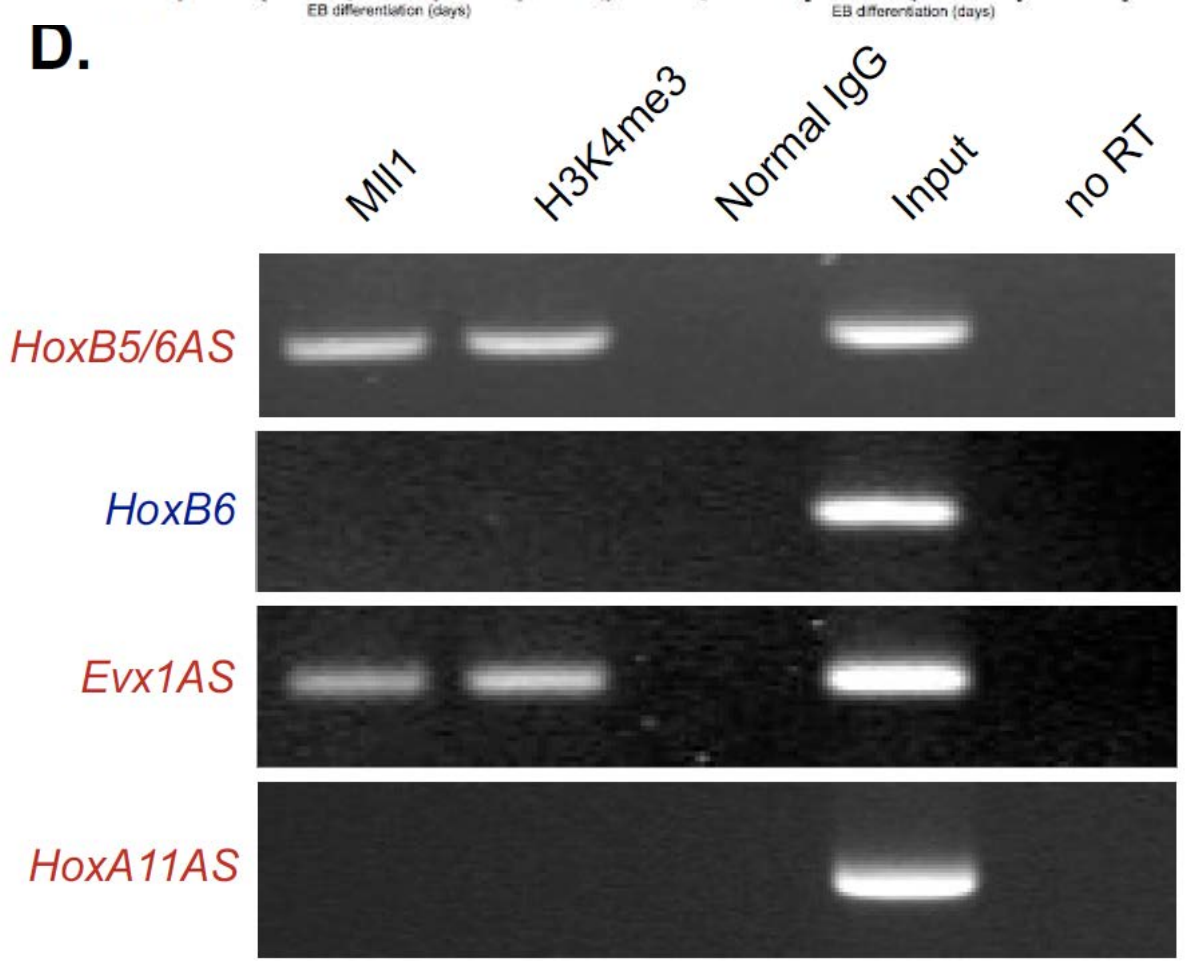
Differentially expressed noncoding transcripts during embryonic stem cell differentiation



Dinger et al. (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research* 18: 1433–1445 (2008).

Dinger *et al.* (2008)
Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation.
Genome Research
18: 1433–1445 (2008)





Differentiation-induced transcripts antisense to developmental genes associate with chromatin modifying complexes and modified histones

Epigenetic processes are central to differentiation and development, long-term responses to environmental variables, and brain function.

Epigenetic memory is embedded in the methylation and hydroxy-methylation of cytosines in DNA and in a wide range of modifications of the histones that package DNA into nucleosomes.

These are catalyzed by a suite of ~100 generic enzymes / chromatin modifying complexes that impose a myriad of different chemical marks at hundreds of thousands, if not millions, of genomic locations in different cells at different stages of differentiation.

What determines the site selectivity of these enzymes?

What determines the positioning of nucleosomes?

What is the molecular basis of epigenome-environmental interactions?

Sciencexpress Report

The *Air* Noncoding RNA Epigenetically Silences Transcription by Targeting G9a to Chromatin

Takashi Nagano,[1,2] Jennifer A. Mitchell,[1] Lionel A. Sanz,[3] Florian M. Pauler,[4] Anne C. Ferguson-Smith,[5] Robert Feil,[3] Peter Fraser[1]*

Mol Cell. 2008 Oct 24;32(2):232-46.

Related Articles, Links

Cell Press

Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C.

Department of Genetics and Pathology, Dag Hammarskjölds Väg 20, Rudbeck Laboratory, Uppsala University, 751 85 Uppsala, Sweden.

Recent investigations have implicated long antisense noncoding RNAs in the epigenetic regulation of chromosomal domains. Here we show that Kcnq1ot1 is an RNA polymerase II-encoded, 91 kb-long, moderately stable nuclear transcript and that its stability is important for bidirectional silencing of genes in the Kcnq1 domain. Kcnq1ot1 interacts with chromatin and with the H3K9- and H3K27-specific histone methyltransferases G9a and the PRC2 complex in a lineage-specific manner. This interaction correlates with the presence of extended regions of chromatin enriched with H3K9me3 and H3K27me3 in the Kcnq1 domain in placenta, whereas fetal liver lacks both chromatin interactions and heterochromatin structures. In addition, the Kcnq1 domain is more often found in contact with the nucleolar compartment in placenta than in liver. Taken together, our data describe a mechanism whereby Kcnq1ot1 establishes lineage-specific transcriptional silencing patterns through recruitment of chromatin remodeling complexes and maintenance of these patterns through subsequent cell divisions occurs via targeting the associated regions to the perinucleolar compartment.

Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression

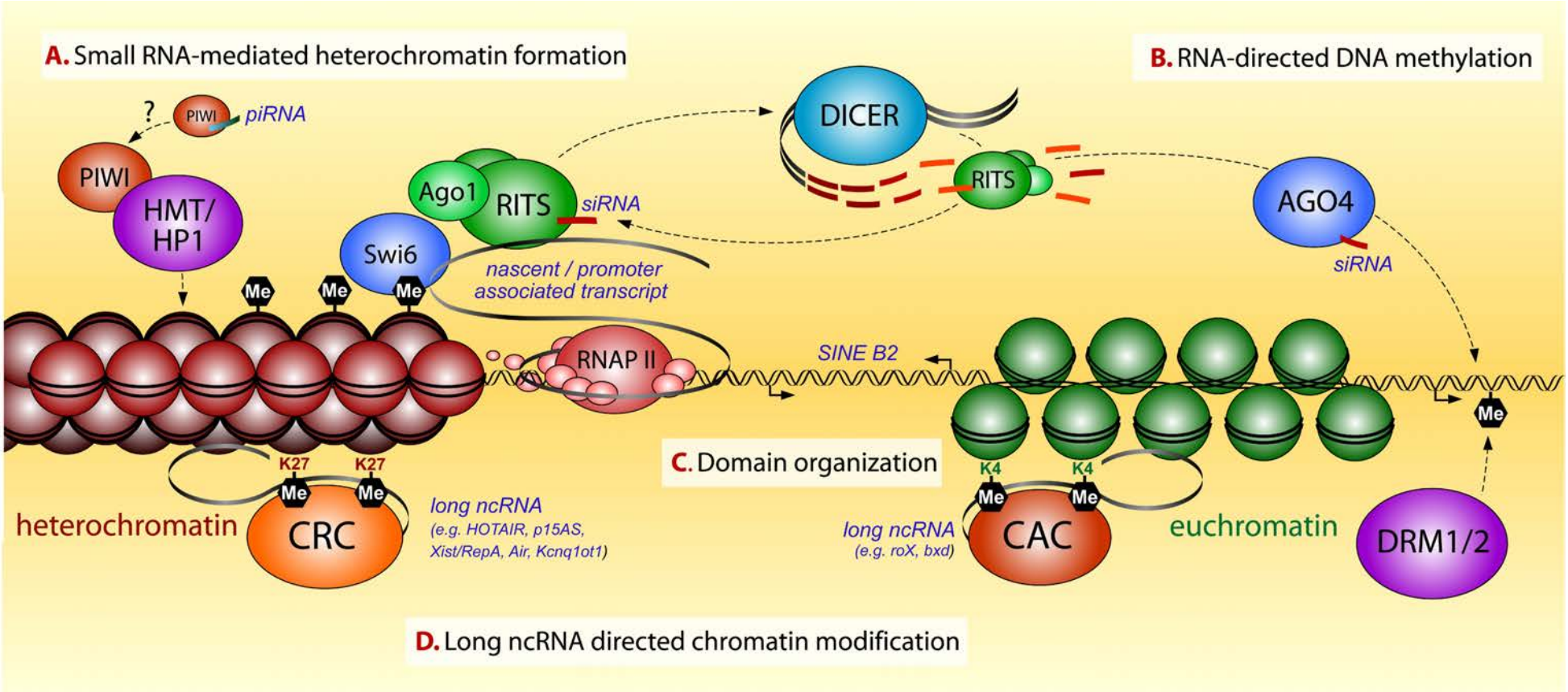
Ahmad M. Khalil[a,b,1], Mitchell Guttman[a,c,1], Maite Huarte[a,b], Manuel Garber[a], Arjun Raj[d], Dianali Rivea Morales[a,b], Kelly Thomas[a,b], Aviva Presser[a], Bradley E. Bernstein[a,e], Alexander van Oudenaarden[d], Aviv Regev[a,c], Eric S. Lander[a,c,f,1,2], and John L. Rinn[a,b,1,2]

[a]The Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142; [b]Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215; Departments of [c]Biology and [d]Physics, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129; and [f]Department of Systems Biology, Harvard Medical School, Boston, MA 02114

Contributed by Eric S. Lander, May 3, 2009 (sent for review March 15, 2009)

We recently showed that the mammalian genome encodes >1,000 large intergenic noncoding (linc)RNAs that are clearly conserved across mammals and, thus, functional. Gene expression patterns have implicated these lincRNAs in diverse biological processes, including cell-cycle regulation, immune surveillance, and embryonic stem cell pluripotency. However, the mechanism by which these lincRNAs function is unknown. Here, we expand the catalog of human lincRNAs to ≈3,300 by analyzing chromatin-state maps of various human cell types. Inspired by the observation that the well-characterized lincRNA HOTAIR binds the polycomb repressive complex (PRC)2, we tested whether many lincRNAs are physically associated with PRC2. Remarkably, we observe that ≈20% of lincRNAs expressed in various cell types are bound by PRC2, and that additional lincRNAs are bound by other chromatin-modifying complexes. Also, we show that siRNA-mediated depletion of certain lincRNAs associated with PRC2 leads to changes in gene expression, and that the up-regulated genes are enriched for those normally silenced by PRC2. We propose a model in which some lincRNAs guide chromatin-modifying complexes to specific genomic loci to regulate gene expression.

intergenic noncoding (linc)RNAs. These lincRNAs show similar expression levels as protein-coding genes, but lack any protein-coding capacity. Importantly, lincRNAs show significant evolutionary conservation relative to neutral sequences, providing strong evidence that they have been functional in the mammalian lineage (1). We note that nonconserved RNA sequences identified in other collections could be functional, but biological evidence such as loss-of-function experiments would be needed to establish their functionality (5) (Fig. S1*A*). Previous studies by us and others have demonstrated that groups of lincRNAs exhibit expression patterns across cell types and tissues that correlate with patterns seen for protein-coding genes involved in cellular processes such as cell-cycle regulation, innate immunity responses, and stem cell pluripotency (1, 14). Although these studies clearly demonstrate that there are many functional lincRNAs, key questions remain, including: How many lincRNAs are encoded in mammalian genomes? How do lincRNAs exert their functions? To begin to investigate the number of lincRNAs, we extended our approach of mapping K4-K36 domains to 6 human cell types. The results expand our catalog to



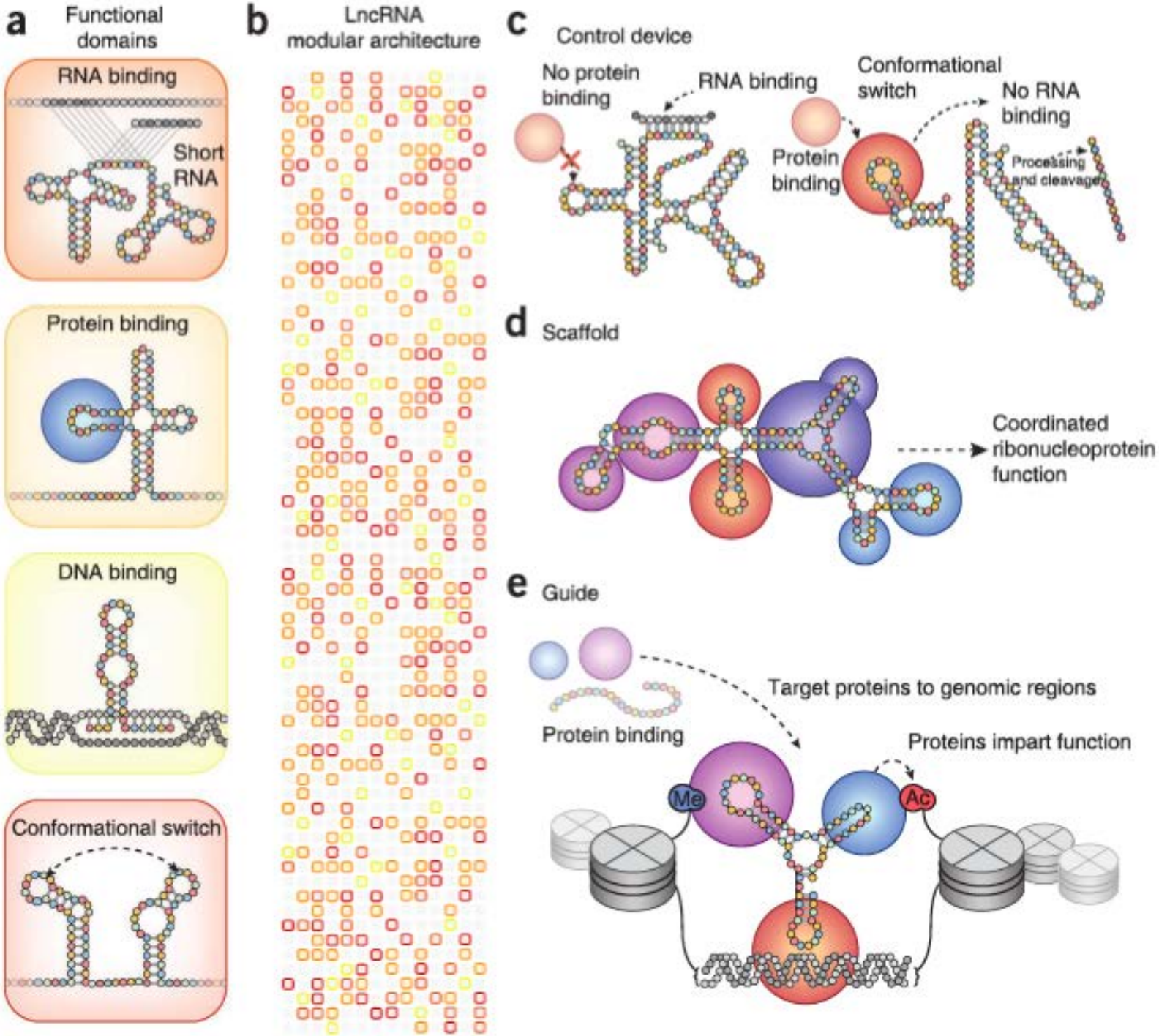
A. Various small RNAs direct chromatin modifications. PIWI proteins and piRNAs interact with HMT/HP1a to induce heterochromatin formation in *Drosophila*. RNA duplexes may be processed in a DICER dependent manner into siRNAs that may subsequently direct chromatin modifications, possibly by targeting nascent transcripts or DNA directly. siRNAs may direct histone methylation (Me) via RITS (RNA-induced transcriptional silencing complex) in centromere heterochromatin in fission yeast.

B. siRNAs originating from RNA Polymerase 4 transcripts can direct DNA methylation in plants.

C. Transcription of SINE B2 elements can establish boundaries between euchromatin and heterochromatin domains in mouse.

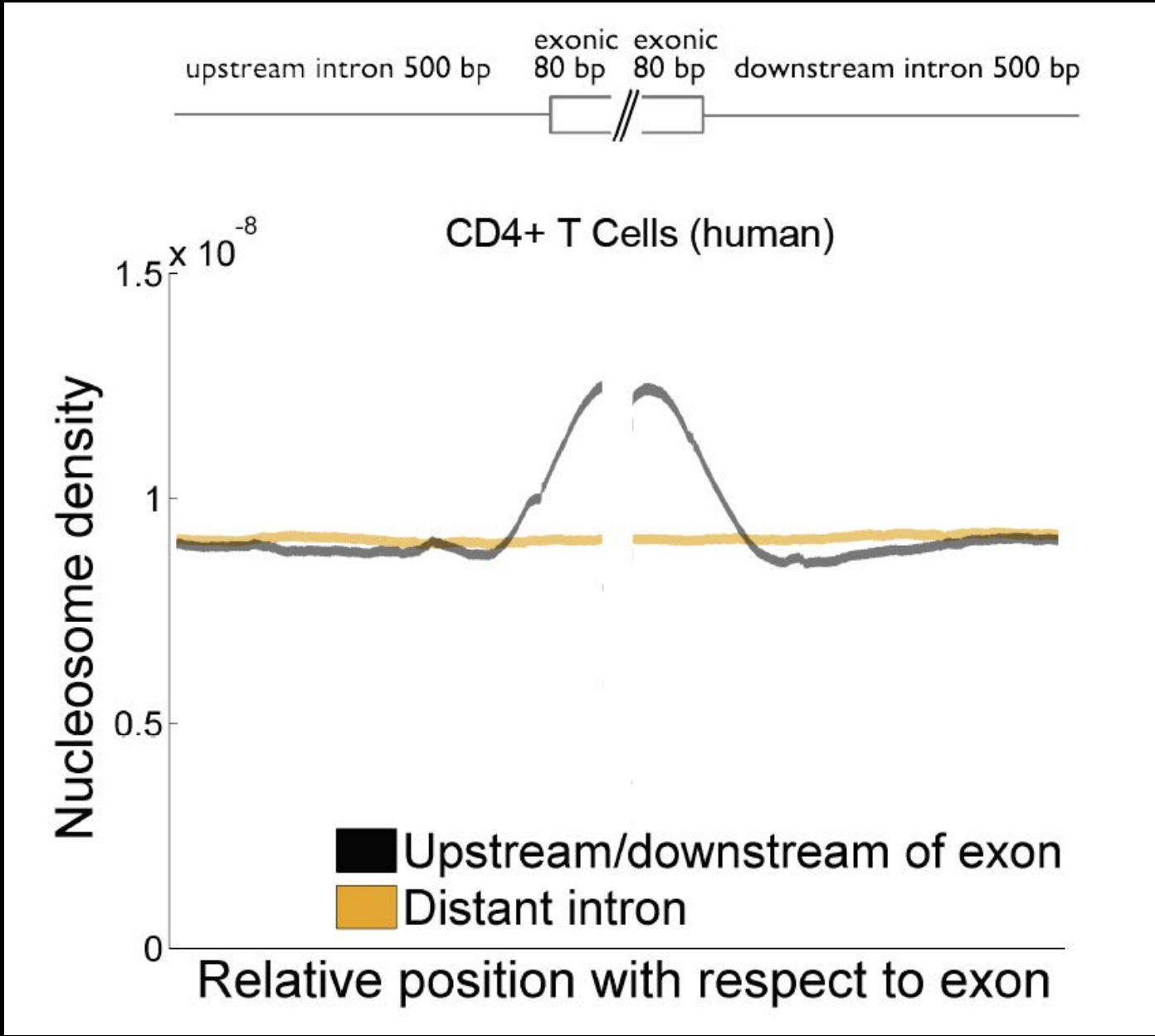
D. Long ncRNAs can recruit chromatin repressor complexes (CRC) or chromatin activating complexes (CAC) to target loci in cis or trans, thereby regulating the chromatin context of local genes.

Mattick et al. (2009) RNA regulation of epigenetic processes. *Bioessays* 31: 51-59.



Mercer & Mattick (2013) *Nature Structural & Molecular Biology* 20: 300-307

Nucleosomes are preferentially positioned at exons in somatic and germ cells in vertebrates



Satu Nahkuri, Ryan Taft and John Mattick (2009) *Cell Cycle* 8: 3420-3424.

Nature Structural and Molecular Biology 17, 1030-1034 (2010).

nature structural & molecular biology

Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans

Ryan J Taft[1,8], Cas Simons[1,2,8], Satu Nahkuri[1], Harald Oey[1], Darren J Korbie[1], Timothy R Mercer[1], Jeff Holst[3,4], William Ritchie[3,4], Justin J-L Wong[3], John E J Rasko[3–5], Daniel S Rokhsar[6], Bernard M Degnan[7] & John S Mattick[1]

We have recently shown that transcription initiation RNAs (tiRNAs) are derived from sequences immediately downstream of transcription start sites. Here, using cytoplasmic and nuclear small RNA high-throughput sequencing datasets, we report the identification of a second class of nuclear-specific ~17- to 18-nucleotide small RNAs whose 3′ ends map precisely to the splice donor site of internal exons in animals. These splice-site RNAs (spliRNAs) are associated with highly expressed genes and show evidence of developmental stage– and region–specific expression. We also show that tiRNAs are localized to the nucleus, are enriched at chromatin marks associated with transcription initiation and possess a 3′-nucleotide bias. Additionally, we find that microRNA-offset RNAs (moRNAs), the *miR-15/16* cluster previously linked to oncosuppression and most small nucleolar RNA (snoRNA)-derived small RNAs (sdRNAs) are enriched in the nucleus, whereas most miRNAs and two H/ACA sdRNAs are cytoplasmically enriched. We propose that nuclear-localized tiny RNAs are involved in the epigenetic regulation of gene expression.

RNA as the substrate for epigenome-environment interactions

RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning, and cognition

*John S. Mattick**

Introduction

Animal development and neurological function are critically dependent on inbuilt and environmentally influenced epigenetic processes that alter chromatin structure and hence gene expression patterns at many loci around the genome. Here I consider the implications of the increasing evidence that RNA directs chromatin-modifying complexes to their sites of action, and that RNA is widely edited, especially in the brain. Editing capacity and activity have expanded during vertebrate, mammalian and primate evolution, wherein the majority targets noncoding sequences, many of which are derived from retrotransposed elements. Heuristically joining these dots leads to the obvious possibility that RNA editing alters regulatory circuitry and can feedback into epigenetic memory, and that the expansion of the enzymatic repertoire for RNA editing along with mobilizable target cassettes was central to the emergence of phenotypic plasticity, learning, and cognition. It also suggests that the widespread colonization of mammalian genomes by transposable elements and the pervasive differential transcription of noncoding sequences are not due to selfish elements and noisy transcription, as often thought, but to an evolved capacity that harnessed RNA and retrotransposons as plastic substrates, underpinning phenotypic adaptability and information storage. Finally, the multiple parallels between the nervous and immune systems suggests that they use similar processes, many of which are RNA-related, to induce somatic plasticity and fine scale specificity, especially in intercellular and intermolecular recognition.

Gene-environment interactions and epigenetic memory

Gene-environment interactions occur at two levels. Short-term responses to physiological variables are largely transduced by signal transduction cascades that alter gene expression.

RNA editing

Two types, both involve base deamination:

A > I - catalyzed by ADARs ("Adenosine Deaminases that Act on RNA"

ADAR1 and ADAR2 occur in most animals, are expressed in most tissues but highly expressed in brain / nervous system. Developmentally lethal.

ADAR 3 is vertebrate-specific and brain-specific. Function unknown.

C / 5meC > U / T - catalyzed by APOBECs ("ApoB Editing Complex")

5 families of APOBECs, 3 vertebrate-specific, 2 mammal-specific.

The APOBEC3 family expands from one ortholog in mouse to 8 in human (APOBEC3A-H), with very strong signatures of positive selection.

Table 1: ADAR substrates with editing sites in coding sequence.

RNA	Base-pairing	Codon changes[a]	Functional changes	Reference
Glutamate receptor				
gluR-B (AMPA)	Exon/intron	Q/R^{607}, R/G^{764}	Editing of Q/R site lowers Ca^{2+} permeability, & receptor tends to be retained at ER as monomer	[4, 8]
gluR-C (AMPA)	Exon/intron	R/G^{769}	Non-edited GluR-$B^{Q/Q}$ causes epileptic seizures and death within 3 weeks of birth	[8]
gluR-D (AMPA)	Exon/intron	R/G^{765}	Editing of R/G site enhances recovery from desensitization	[8]
gluR-5 (kainate)	Exon/intron	Q/R^{636}	Editing of Q/R site in kainate receptors potentiates inhibition of receptors by membrane fatty acids	[4]
gluR-6 (kainate)	Exon/intron	Q/R^{621}, I/V^{567}, Y/C^{571}	Editing of all 3 sites in gluR-6 increases higher Ca^{2+} permeability	[4, 7]
K_v1.1 channel	Exon/exon	I/V^{400}	Rapid recovery from inactivation, shortening duration of and increasing frequency of action potential	[13, 15]
$GABA_A$-α3 receptor	Exon/intron	I/M^{342}	Smaller peak current amplitudes, slower activation, and faster deactivation compared to non edited receptors	[16, 17]
Serotonin receptor	Exon/intron	$I/V^{157\ \&\ 161}$, I/M^{157}, N/D^{159}, N/S^{159}, N/G^{159}	Lower coupling efficacy to G-protein Lower tendency to isomerizes, hence lower level of constitutive activity	[20, 21]

[a] *Editing sites are named according to the amino acid change they produced and amino acid position (unedited/edited [amino acid position])*

nature biotechnology

Systematic identification of abundant A-to-I editing sites in the human transcriptome

Open access, freely available online PLoS BIOLOGY

Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome

Alekos Athanasiadis[1,2], Alexander Rich[2], Stefan Maas[1*]

1 Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania, United States of America, 2 Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

Editing is not restricted to a neuroreceptor mRNAs but occurs in thousands of transcripts.

RNA editing by members of the ADAR (adenosine deaminases acting on RNA) family leads to site-specific conversion of

editing in the human transcriptome. An experimental demonstration in 43 genes was extended by a broader computational analysis of more than 100,000 human mRNAs. We find that 1,445 human mRNAs (1.4%) are subject to

Editing occurs mainly in noncoding sequences, implying that editing is altering regulatory circuits and networks, potentially influencing RNA-directed epigenetic memory.

the role of editing in controlling dsRNA stability.

Citation: Athanasiadis A, Rich A, Maas S (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol 2(12): e391.

There is a massive increase in the amount and intensity of A>I editing of human RNAs compared to mouse (35x increase).

Widespread RNA Editing of Embedded *Alu* Elements in the Human Transcriptome

A survey of RNA editing in human brain

The vast majority of this increase occurs in Alu sequences, primate-specific SINEs that invaded in three waves during primate evolution and occupy 10.5% of the human genome (~1.2 million largely sequence-unique copies).

75% of all known genes having *Alu* insertions within their introns and/or UTRs. Transcribed *Alu* sequences can alter splicing patterns by generating new exons, but other impacts of intragenic *Alu* elements on their host RNA are largely unexplored. Recently, repeat elements present in the introns or 3′-UTRs of 15 human brain RNAs have been shown to be targets for multiple adenosine to inosine (A-to-I) editing. Using a statistical approach, we find that editing of transcripts with embedded *Alu* sequences is a global phenomenon in the human transcriptome, observed in 2674 (~2%) of all publicly available full-length human cDNAs ($n = 128{,}406$), from >250 libraries and >30 tissue sources. In the vast majority of edited RNAs, A-to-I substitutions are clustered within transcribed sense or antisense *Alu* sequences. Edited bases are primarily associated with retained introns, extended UTRs, or with transcripts that have no corresponding known gene. Therefore, *Alu*-associated RNA editing may be a mechanism for marking nonstandard transcripts, not destined for translation.

in translated exons and few in untranslated exons. Most edits were in high-copy-number repeats, usually *Alu*s. Analysis of the genome in the vicinity of edited sequences strongly supports the idea that formation of intramolecular double-stranded RNA with an inverted copy underlies most A→I editing. The likelihood of editing is increased by the presence of two inverted copies of a sequence within the same intron, proximity of the two sequences to each other (preferably within 2 kb), and by a high density of inverted copies in the vicinity. Editing exhibits sequence preferences and is less likely at an adenosine 3′ to a guanosine and more likely at an adenosine 5′ to a guanosine. Simulation by BLAST alignment of the double-stranded RNA molecules that underlie known edits indicates that there is a greater likelihood of A→I editing at A:C mismatches than editing at other mismatches or at A:U matches. However, because A:U matches in double-stranded RNA are more common than all mismatches, overall the likely effect of editing is to increase the number of mismatches in double-stranded RNA.

Genome Res. 2004 14: 1719-1725

Genome Res. 2004 14: 2379-2387

A-to-I RNA editing shapes transcriptome diversity in primates

Nurit Paz-Yaacov[a,b], Erez Y. Levanon[c], Eviatar Nevo[d,1], Yaron Kinar[e], Alon Harmelin[f], Jasmine Jacob-Hirsch[a], Ninette Amariglio[a], Eli Eisenberg[g], and Gideon Rechavi[a,b,1]

[a]Cancer Research Center, Chaim Sheba Medical Center, Tel Hashomer 52621, Israel; [b]Sackler School of Medicine and [g]Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel; [c]Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, Ramat Gan 52900, Israel; [d]Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel; [e]Compugen Ltd., Tel Aviv 69512, Israel; and [f]Department of Veterinary Resources, Weizmann Institute of Science, Rehovot 76100, Israel

The extent and intensity of A>I editing also increases during primate evolution

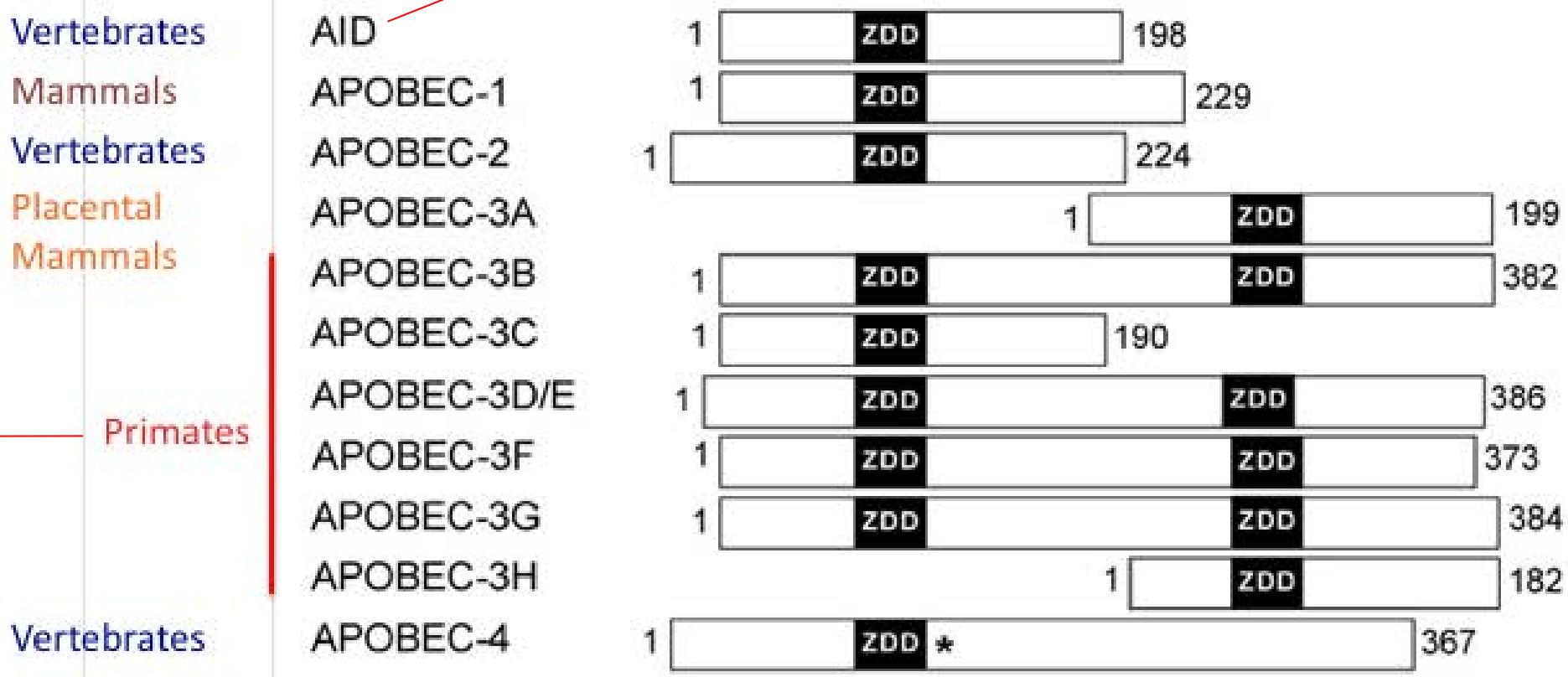
Human and chimpanzee genomes are almost identical, yet humans express higher brain capabilities. Deciphering the basis for this superiority is a long sought-after challenge. Adenosine-to-inosine (A-to-I) RNA editing is a widespread modification of the transcriptome. The editing level in humans is significantly higher compared with nonprimates, due to exceptional editing within the primate-specific *Alu* sequences, but the global editing level of nonhuman primates has not been studied so far. Here we report the sequencing of transcribed *Alu* sequences in humans, chimpanzees, and rhesus monkeys. We found that, on average, editing level in the transcripts analyzed is higher in human brain compared with nonhuman primates, even where the genomic *Alu* structure is unmodified. Correlated editing is observed for pairs and triplets of specific adenosines along the *Alu* sequences. Moreover, new editable species-specific *Alu* insertions, subsequent to the human–chimpanzee split, are significantly enriched in genes related to neuronal functions and neurological diseases. The enhanced editing level in the human brain and the association with neuronal functions both hint at the possible contribution of A-to-I editing to the development of higher brain function. We show here that combinatorial editing is the most significant contributor to the transcriptome repertoire and suggest that *Alu* editing adapted by natural selection may therefore serve as an alternate information mechanism based on the binary A/I code.

RNA-specific adenosine deaminase acting on RNA (ADAR) enzyme family and appears to be tissue specific with brain tissue being the most edited (13–16). The splicing and translational machineries recognize inosine (I) as guanosine (G). Therefore, the result of ADAR-mediated editing consists of genomically encoded adenosines that are read as guanosines in the RNA sequence. Many of the RNA editing targets play a central role in neurogenesis. Indeed, disruption of the editing process in lower organisms such as *Caenorhabditis elegans* and *Drosophila melanogaster* resulted in behavioral and neural defects (17, 18). Moreover altered editing patterns in humans and mice have been linked mainly to neuropathological disorders, such as amyotrophic lateral sclerosis, epilepsy, and brain tumors (19–23).

RNA editing in humans occurs predominantly within the primate-specific *Alu* repetitive elements, affecting thousands of genes in tens of thousands of sites. The overwhelming majority of these sites are located in noncoding sequences (introns and UTRs) (24–27). Still, some A-to-I RNA editing occurs in coding sequences and alters the mature protein sequence and its properties. Interestingly, the level of RNA editing in humans is more than an order of magnitude higher than that in the mouse, rat, chicken, and fly (26, 28). This difference is explained by the dominance of the primate-specific *Alu* elements in the human transcriptome, which can generate double-stranded RNA (dsRNA) structures (29). How-

PNAS proof
Embargoed

Involved in somatic rearrangement and hypermutation of immunoglobulin domains in B-cells and T-cells



APOBEC3F and 3G appear to control exogenous and endogenous retroviral and LINE-1 retrotransposition.

APOBEC3G is expressed in post-mitotic neurons.

doi:10.1038/nature08248

LETTERS

L1 retrotransposition in human neural progenitor cells

Nicole G. Coufal[1], José L. Garcia-Perez[2,3], Grace E. Peng[1], Gene W. Yeo[1]†, Yangling Mu[1], Michael T. Lovci[1]†, Maria Morell[4], K. Sue O'Shea[4], John V. Moran[2,5] & Fred H. Gage[1]

Long interspersed element 1 (LINE-1 or L1) retrotransposons have markedly affected the human genome. L1s must retrotranspose in the germ line or during early development to ensure their evolutionary success, yet the extent to which this process affects somatic cells is poorly understood. We previously demonstrated that engineered human L1s can retrotranspose in adult rat hippocampus progenitor cells *in vitro* and in the mouse brain *in vivo*[1]. Here we demonstrate that neural progenitor cells isolated from human fetal brain and derived from human embryonic stem cells support the retrotransposition of engineered human L1s *in vitro*. Furthermore, we developed a quantitative multiplex polymerase chain reaction that detected an increase in the copy number of endogenous L1s in the hippocampus, and in several regions of adult human brains, when compared to the copy number of endogenous L1s in heart or liver genomic DNAs from the same donor. These data suggest that *de novo* L1 retrotransposition events may occur in the human brain and, in principle, have the potential to contribute to individual somatic mosaicism.

To determine whether L1 retrotransposition occurred in undifferentiated cells, we conducted immunocytochemical localization of cell-type-restricted markers in EGFP-positive hCNS-SCns. These cells expressed neural stem cell markers, including SOX2, Nestin, Musashi-1 and SOX1 (Fig. 1e and Supplementary Fig. 2a, b), and some co-labelled with Ki-67, indicating that they continued to proliferate (Supplementary Fig. 2c). EGFP-positive hCNS-SCns could also be differentiated to cells of both the neuronal and the glial lineages (Fig. 1f, g). Notably, $L1_{RP}$ did not retrotranspose using our experimental conditions in primary human astrocytes or fibroblasts, although a low level of endogenous L1 expression was detected in both cell types (Fig. 1d and Supplementary Figs 2d, e and 6a, b).

We next used two different protocols to derive NPCs from five human embryonic stem cell lines (hESCs; Fig. 2a). As in our previous study[1], NPC differentiation led to a ~25-fold increase in L1 promoter activity over a 2-day period, and then a decline (Fig. 2c); there was also a ~250-fold increase in synapsin promoter activity during differentiation (Supplementary Fig. 4b). H13B-derived NPCs expressed both

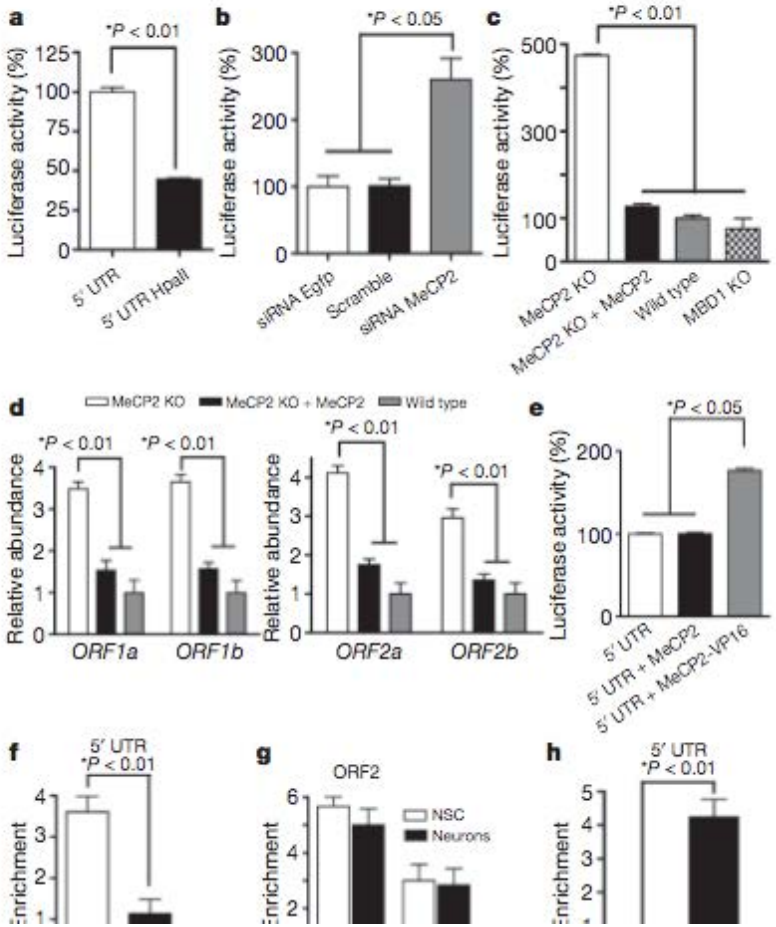
doi:10.1038/nature09544

L1 retrotransposition in neurons is modulated by MeCP2

Alysson R. Muotri[1]*, Maria C. N. Marchetto[2]*, Nicole G. Coufal[2], Ruth Oefner[2], Gene Yeo[3], Kinichi Nakashima[4] & Fred H. Gage[2]

Long interspersed nuclear elements-1 (LINE-1 or L1s) are abundant retrotransposons that comprise approximately 20% of mammalian genomes[1–3]. Active L1 retrotransposons can impact the genome in a variety of ways, creating insertions, deletions, new splice sites or gene expression fine-tuning[4–6]. We have shown previously that L1 retrotransposons are capable of mobilization in neuronal progenitor cells from rodents and humans and evidence of massive L1 insertions was observed in adult brain tissues but not in other somatic tissues[7,8]. In addition, L1 mobility in the adult hippocampus can be influenced by the environment[9]. The neuronal specificity of somatic L1 retrotransposition in neural progenitors is partially due to the transition of a Sox2/HDAC1 repressor complex to a Wnt-mediated T-cell factor/lymphoid enhancer factor (TCF/LEF) transcriptional activator[7,10]. The transcriptional switch accompanies chromatin remodelling during neuronal differentiation, allowing a transient stimulation of L1 transcription[7]. The activity of L1 retrotransposons during brain development can have an impact on gene expression and neuronal function, thereby increasing brain-specific genetic mosaicism[11,12]. Further understanding of the molecular mechanisms that regulate L1 expression should provide new insights into the role of L1 retrotransposition during brain development. Here we show that L1 neuronal transcription and retrotransposition in rodents are increased in the absence of methyl-CpG-binding protein 2 (MeCP2), a protein involved in global DNA methylation and human neurodevelopmental diseases. Using neuronal progenitor cells derived from human induced pluripotent stem cells and human tissues, we revealed that patients with Rett syndrome (RTT), carrying MeCP2 mutations, have increased susceptibility for L1 retrotransposition. Our data demonstrate that L1 retrotransposition can be controlled in a tissue-specific manner and that disease-related genetic mutations can influence the frequency of neuronal L1 retrotransposition. Our findings add a new level of complexity to the molecular events that can lead to neurological disorders.

We repeated the luciferase assay using neuroepithelial cells from a sibling MBD1 KO animal[16]. MBD1 (methyl-CpG binding domain protein 1) is part of the methyl-binding protein family and has differential



LETTER

doi:10.1038/nature10531

Somatic retrotransposition alters the genetic landscape of the human brain

J. Kenneth Baillie[1]*, Mark W. Barnett[1]*, Kyle R. Upton[1]*, Daniel J. Gerhardt[2], Todd A. Richmond[2], Fioravante De Sapio[1], Paul Brennan[3], Patrizia Rizzu[4], Sarah Smith[1], Mark Fell[1], Richard T. Talbot[1], Stefano Gustincich[5], Thomas C. Freeman[1], John S. Mattick[6], David A. Hume[1], Peter Heutink[4], Piero Carninci[7], Jeffrey A. Jeddeloh[2] & Geoffrey J. Faulkner[1]

Retrotransposons are mobile genetic elements that use a germline 'copy-and-paste' mechanism to spread throughout metazoan genomes[1]. At least 50 per cent of the human genome is derived from retrotransposons, with three active families (L1, *Alu* and SVA) associated with insertional mutagenesis and disease[2,3]. Epigenetic and post-transcriptional suppression block retrotransposition in somatic cells[4,5], excluding early embryo development and some malignancies[6,7]. Recent reports of L1 expression[8,9] and copy number variation[10,11] in the human brain suggest that L1 mobilization may also occur during later development. However, the corresponding integration sites have not been mapped. Here we apply a high-throughput method to identify numerous L1, *Alu* and SVA germline mutations, as well as 7,743 putative somatic L1 insertions, in the hippocampus and caudate nucleus of three individuals. Surprisingly, we also found 13,692 somatic *Alu* insertions and 1,350 SVA insertions. Our results demonstrate that retrotransposons mobilize to protein-coding genes differentially expressed and active in the brain. Thus, somatic genome mosaicism driven by retrotransposition may reshape the genetic circuitry that underpins normal and abnormal neurobiological processes.

Mapping the individual retrotransposition events that collectively form a somatic mosaic is challenging owing to the rarity of each mutant allele in a heterogeneous cell population. We therefore developed a high-throughput protocol that we call retrotransposon capture sequencing (RC-seq). First, fragmented genomic DNA was hybridized to custom sequence capture arrays targeting the 5′ and 3′ termini of full-length L1, *Alu* and SVA retrotransposons (Fig. 1a and Supplementary Tables 1 and 2). Immobile ERVK and ERV1 long terminal repeat (LTR) elements were included as negative controls. Second, the captured DNA was deeply sequenced, yielding ~25 million paired-end 101-mer reads per sample (Fig. 1b). Last, read pairs were mapped using a conservative computational pipeline designed to identify known (Fig. 1c) and novel (Fig. 1d and Supplementary Fig. 1a–d) retrotransposon insertions with uniquely mapped read pairs ('diagnostic reads') spanning their termini.

Previous works have equated L1 CNV with somatic mobilization *in vivo*[10,11]. To test this assumption with RC-seq, we first screened five brain subregions taken from three individuals (donors A, B and C) for L1 CNV. A significant ($P < 0.001$) increase was observed in the number of copies of L1 open reading frame 2 (ORF2) present in DNA

PERSPECTIVE

The Eukaryotic Genome as an RNA Machine

Paulo P. Amaral, Marcel E. Dinger, Tim R. Mercer, John S. Mattick*

The past few years have revealed that the genomes of all studied eukaryotes are almost entirely transcribed, generating an enormous number of non–protein-coding RNAs (ncRNAs). In parallel, it is increasingly evident that many of these RNAs have regulatory functions. Here, we highlight recent advances that illustrate the diversity of ncRNA control of genome dynamics, cell biology, and developmental programming.

RNAs are an integral component of chromosomes and contribute to their structural organization (*1*, *2*). It is now becoming apparent that chromatin architecture and epigenetic memory are regulated by RNA-directed processes that, although the exact mechanisms are yet to be understood, involve the recruitment of histone-modifying complexes and DNA methyltransferases to specific loci (*3*). Whereas long non–protein-coding RNAs (ncRNAs) have been classically implicated in the regulation of dosage compensation and genomic imprinting in animals (*4*), they seem to play a much broader role in the epigenetic control of developmental trajectories (*5*). For example, it was recently shown that 231 long ncRNAs associated with human *HOX* gene clusters are colinearly expressed along developmental axes (*5*), one of which, termed *HOTAIR*, transcribed from the *HOXC* locus, was studied in detail and found to recruit Polycomb complexes to repress gene expression in trans at the *HOXD* cluster (*5*) (Fig. 1). Other ncRNAs will likely perform similar functions, such as the intergenic transcripts from globin and antigen receptor loci, which have been associated with complex epigenetic phenomena (*6*, *7*).

Small ncRNAs have been consistently linked with heterochromatin formation via the RNA interference (RNAi) pathway (*8*), including Piwi-interacting RNAs (piRNAs) (*9*), which guide PIWI family proteins to control transposon activity from flies to vertebrates (*10*). However, piRNAs might also regulate euchromatin formation, given that PIWI is required for establishing euchromatin in certain subtelomeric regions in *Drosophila* (*11*).

Higher-level nuclear organization and chromosome dynamics are also regulated by ncRNAs in a variety of systems. For example, the formation of the kinetochore and centromeric heterochromatin in fission yeast is dependent on cell cycle–regulated centromeric repeat–derived RNAs and the RNAi pathway, whereas kinetochore assembly and chromosome segregation require the ribonuclease activity of a component of the exosome (*12*–*15*). These findings reveal an RNA-based mechanistic link between these processes in mitosis. In *Tetrahymena*, RNAs direct heterochromatin formation and DNA elimination via RNAi-dependent recruitment of Polycomb complexes and histone methylation (*16*). The RNAi pathway along with directed histone modifications also regulates the organization of the nucleolus in *Drosophila* (*17*).

Likewise, long ncRNAs direct programmed whole-genome rearrangements during ciliate differentiation (*18*). In mammals, transcription of long ncRNAs contributes to various processes including T cell receptor recombination (*7*), maintenance of telomeres (*19*, *20*), X-chromosome pairing required for dosage compensation (*21*) and inactive X-chromosome perinucleolar localization (*22*).

The functional organization of chromatin can also be regulated by ncRNAs derived from repetitive elements. In mice, bidirectional transcription of a retrotransposed SINE B2 sequence by RNA polymerase (RNAP)II and RNAPIII relocates the associated growth hormone locus into nuclear compartments and locally defines the heterochromatin-euchromatin boundary, regulating the expression of the gene during organogenesis (*23*) (Fig. 1). Given the abundance of transcribed repetitive sequences, this may represent a genome-wide strategy for the control of chromatin domains that may be conserved throughout eukaryotes (*23*–*25*). Moreover, such observations and others suggest that a large portion of the genome may, in fact, be functionally active and that transposon-derived sequences may not be reliable indices of the rate of neutral evolution (*26*).

Transcription

Noncoding RNAs can regulate transcription by interacting with transcription factors, RNAP, or

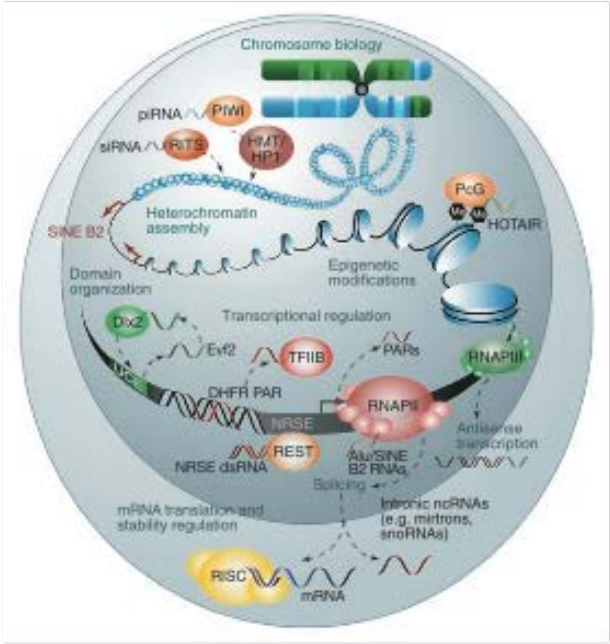
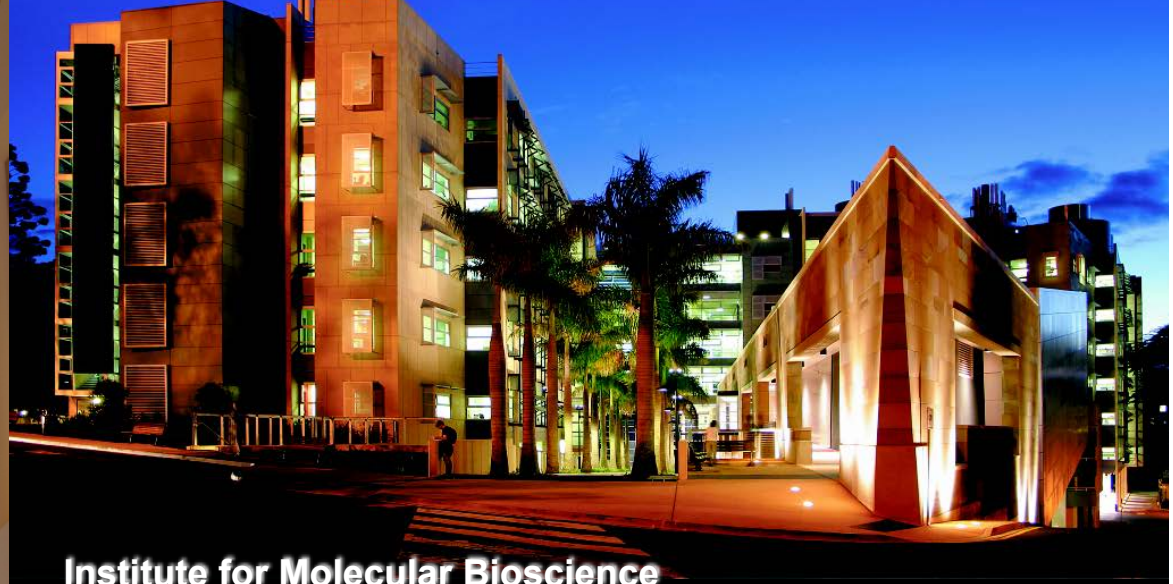


Fig. 1. Recent examples of the various levels of regulation of eukaryotic gene expression and cell biology by ncRNAs. dsRNA, double-stranded RNA; HMT, histone methyltransferases; HP1, heterochromatin protein 1; PARs, promoter-associated RNAs; PcG, Polycomb group proteins; RISC, RNA-induced silencing complex; RITS, RNA-induced initiation of transcriptional gene silencing; siRNA, small interfering RNA; TFIIB, transcription factor IIB; and UCE, ultraconserved element. See text for details, other acronyms, and references. For additional examples, see (*3*, *58*).

Institute for Molecular Bioscience, University of Queensland, St. Lucia QLD 4072, Australia.

*To whom correspondence should be addressed. E-mail: j.mattick@imb.uq.edu.au

Downloaded from www.sciencemag.org on March 27, 2008



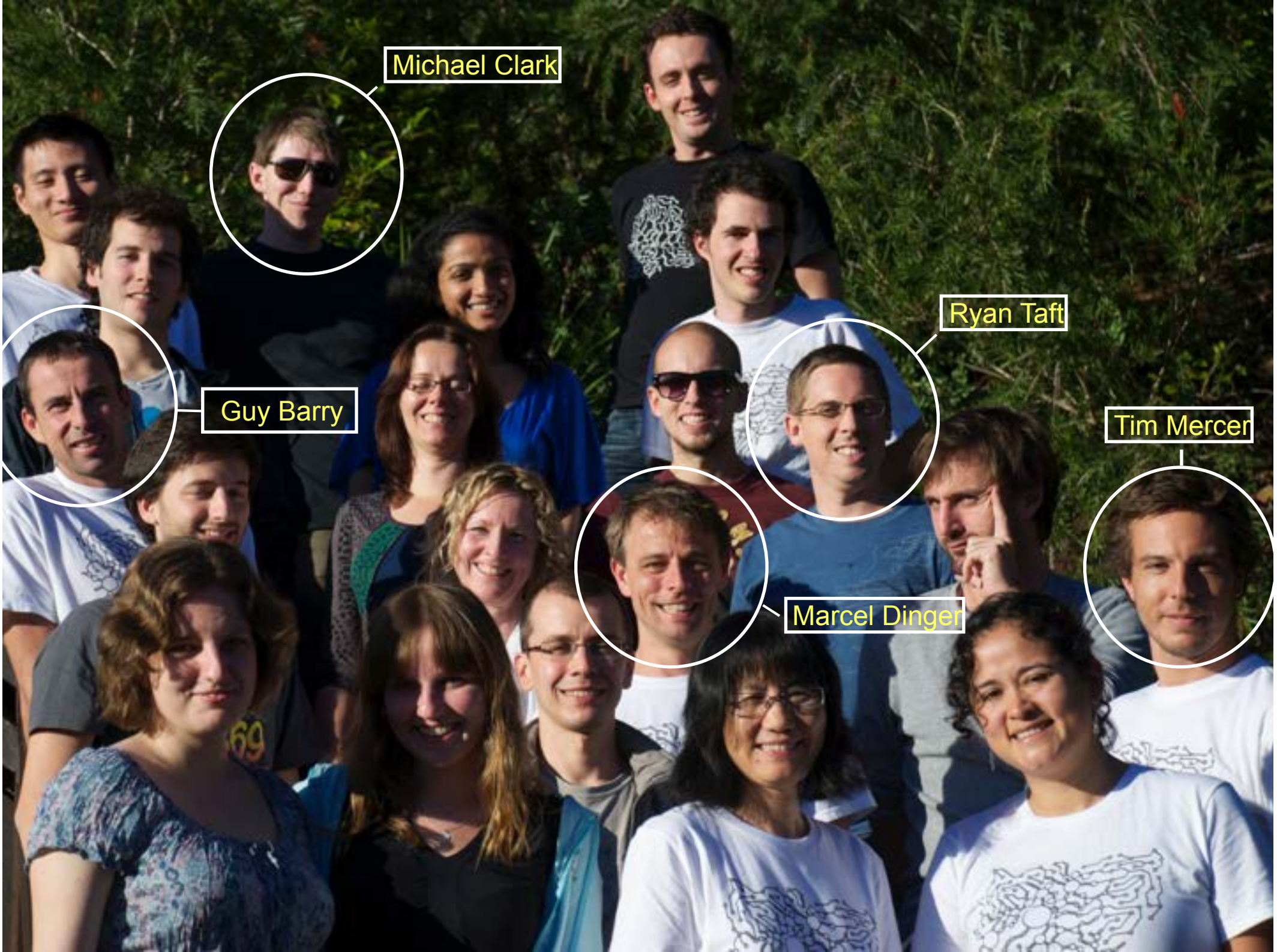
Institute for Molecular Bioscience
University of Queensland
Marjan Askarian-Amiri, Paulo Amaral, Seth Cheetham, Joanne Crawford, Kelin Ru, Ryan Taft

Garvan Institute of Medical Research

Guy Barry, Michael Clark, Marcel Dinger, Tim Mercer, Martin Smith

RIKEN Genomic Science Center
Yoshihide Hayashizaki, Piero Carninci, Harukazu Suzuki, Shintaro Katayama

Allen Brain Institute	Susan Sunkin
Burnham, Florida	Ranjan Perrera
Harvard University	John Rinn
Roche / NimbleGen	Jeff Jeddeloh
Johns Hopkins	Seth Blackshaw, Erin Poth
RIKEN, Wako	Shinichi Nakagawa
AIBN, University of Queensland	Ernst Wolvertang


Michael Clark
Ryan Taft
Guy Barry
Tim Mercer
Marcel Dinger