JC

RK

~~2014-04-03~~

~~2014-04-09~~

2014-04-17

a JC of two (or three) halves

1. **where's miRNA?**
   - recap of miRNA/mRNA binding from AGO IP assays

2. **one transcript to rule them all**
   - *transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene*

3. **neurogenesis in the adult human brain**
   - if there's time…

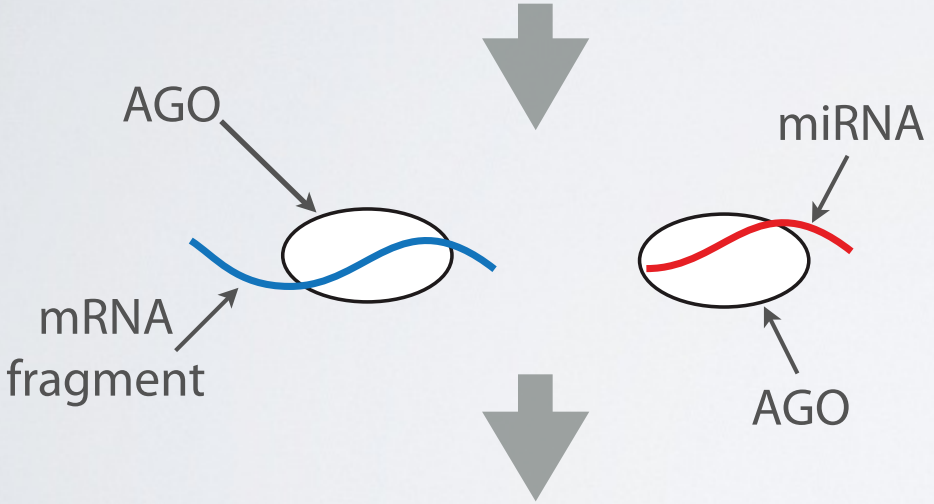
AGO HiTS-CLIP

ARTICLES

Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps

Sung Wook Chi[1], Julie B. Zang[1], Aldo Mele[1] & Robert B. Darnell[1]

MicroRNAs (miRNAs) have critical roles in the regulation of gene expression; however, as miRNA activity requires base pairing with only 6–8 nucleotides of messenger RNA, predicting target mRNAs is a major challenge. Recently, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) has identified functional protein–RNA interaction sites. Here we use HITS-CLIP to covalently crosslink native argonaute (Ago, also called Eif2c) protein–RNA complexes in mouse brain. This produced two simultaneous data sets—Ago–miRNA and Ago–mRNA binding sites—that were combined with bioinformatic analysis to identify interaction sites between miRNA and target mRNA. We validated genome-wide interaction maps for miR-124, and generated additional maps for the 20 most abundant miRNAs present in P13 mouse brain. Ago HITS-CLIP provides a general platform for exploring the specificity and range of miRNA action *in vivo*, and identifies precise sequences for targeting clinically relevant miRNA–mRNA interactions.



search mRNA fragments for potential miRNA binding sites

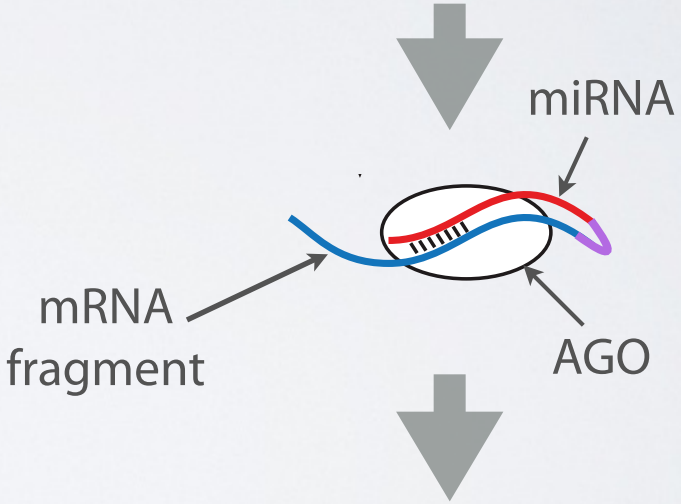
CLASH

Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding

Aleksandra Helwak,[1,2] Grzegorz Kudla,[1,2,3] Tatiana Dudnakova,[1] and David Tollervey[1,*]
[1]Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh EH9 3JR, UK
[2]These authors contributed equally to this work
[3]Present address: MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh EH4 2XU, UK
*Correspondence: d.tollervey@ed.ac.uk
http://dx.doi.org/10.1016/j.cell.2013.03.043

SUMMARY

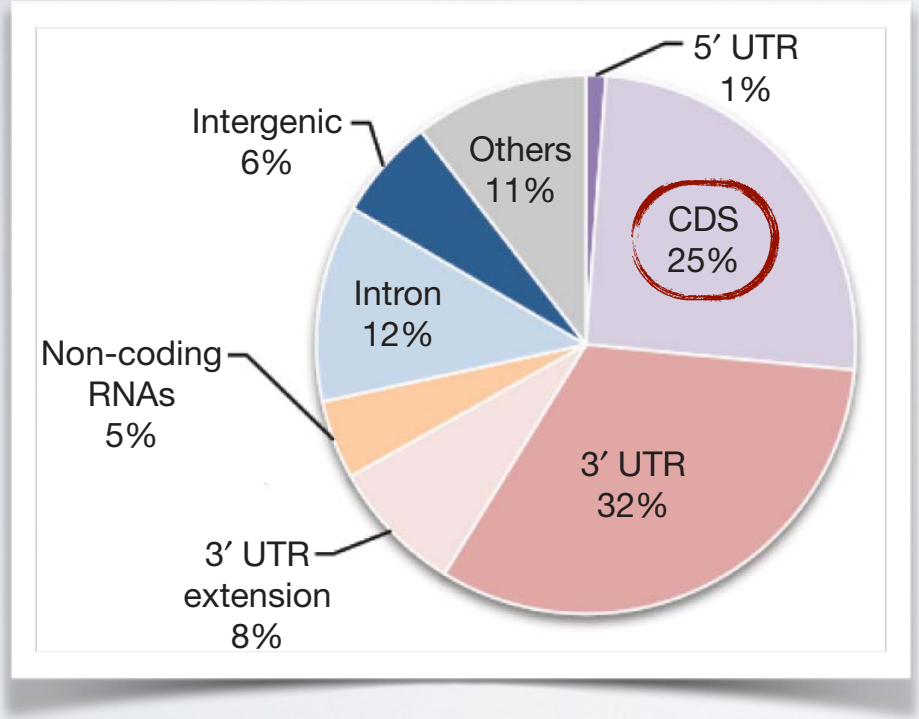
MicroRNAs (miRNAs) play key roles in gene regulation, but reliable bioinformatic or experimental identification of their targets remains difficult. To provide an unbiased view of human miRNA targets, we develpairs with the target; (2) nucleotides paired outside the seed region stabilize interactions but are reported not to influence miRNA efficacy (Garcia et al., 2011; Grimson et al., 2007); and (3) functional miRNA targets are localized close to the extremes of the 3′ UTRs of protein-coding genes in relatively unstructured regions (Grimson et al., 2007). Recently,



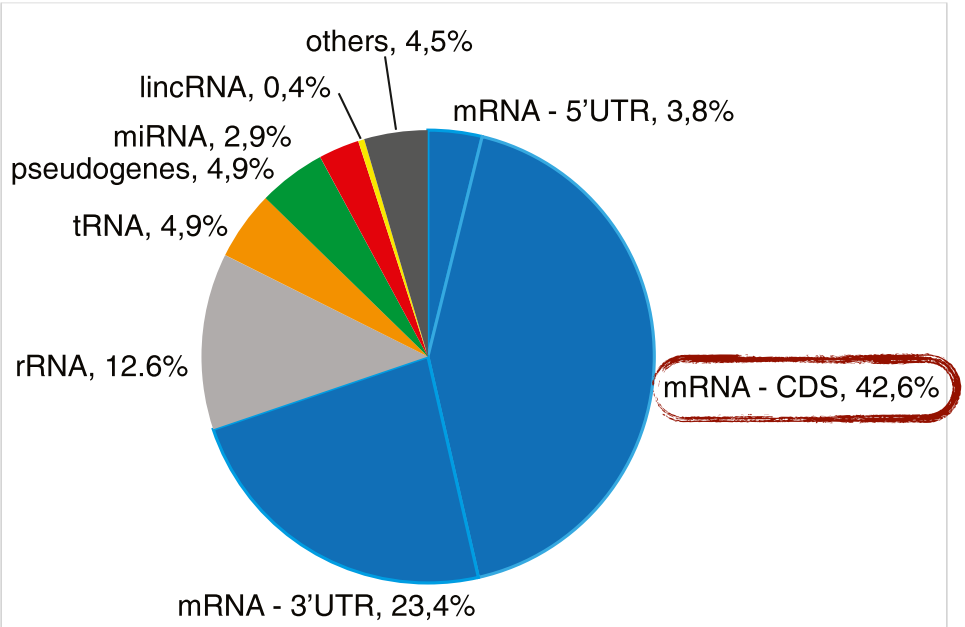
define 1-to-1 targets using miRNA/mRNA chimeric reads

mRNA binding sites

AGO HITS-CLIP



CLASH



- not all miRNA binding sites are in the 3'UTR!

comparing predicted targets

		matches with CLASH	matches with control	enrichment
	Number of interactions	*6,248*	*6,248*	
miRanda	*687,208*	411	29	14.2 ×
PicTar	*205,263*	224	9	24.9 ×
PITA	*192,255*	195	2	97.5 ×
RNAhybrid	*992,584*	310	25	12.4 ×
TargetScan	*54,199*	170	5	34.0 ×
all predictions	*2,131,509*	802	59	13.6 ×

- for all the miRNA binding sites that **are** in the 3'UTR:
  - computational **predictions are enriched** over random
  - however suffer **extremely high false positives & negatives**

a JC of two (or three) halves

1. where's miRNA?
   - recap of miRNA/mRNA binding from AGO IP assays

2. **one transcript to rule them all**
   - transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene

3. **neurogenesis in the adult human brain**
   - if there's time…

Gonzàlez-Porta et al. Genome Biology 2013, 14:R70
http://genomebiology.com/2013/14/7/R70

Genome Biology

RESEARCH Open Access

Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene

Mar Gonzàlez-Porta[1], Adam Frankish[2], Johan Rung[1], Jennifer Harrow[2] and Alvis Brazma[1*]

Abstract

Background: RNA sequencing has opened new avenues for the study of transcriptome composition. Significant evidence has accumulated showing that the human transcriptome contains in excess of a hundred thousand different transcripts. However, it is still not clear to what extent this diversity prevails when considering the relative abundances of different transcripts from the same gene.

Results: Here we show that, in a given condition, most protein coding genes have one major transcript expressed at significantly higher level than others, that in human tissues the major transcripts contribute almost 85 percent to the total mRNA from protein coding loci, and that often the same major transcript is expressed in many tissues. We detect a high degree of overlap between the set of major transcripts and a recently published set of alternatively spliced transcripts that are predicted to be translated utilizing proteomic data. Thus, we hypothesize that although some minor transcripts may play a functional role, the major ones are likely to be the main contributors to the proteome. However, we still detect a non-negligible fraction of protein coding genes for which the major transcript does not code a protein.

Conclusions: Overall, our findings suggest that the transcriptome from protein coding loci is dominated by one transcript per gene and that not all the transcripts that contribute to transcriptome diversity are equally likely to contribute to protein diversity. This observation can help to prioritize candidate targets in proteomics research and to predict the functional impact of the detected changes in variation studies.

Keywords: splicing, transcriptome, gene expression, RNA-seq

analysis summary

- Djebali 2012 suggested genes tend to express a 'major transcript'

- in this paper, the EBI group used 16 **BodyMap** tissues, 5 **ENCODE** cell-lines, and **Flux Simulator** RNA-seq data

- mapped to **Gencode v11** protein coding genes

- transcript quantification using **MISO**, **Cufflinks**, and **MMSEQ**

major transcript

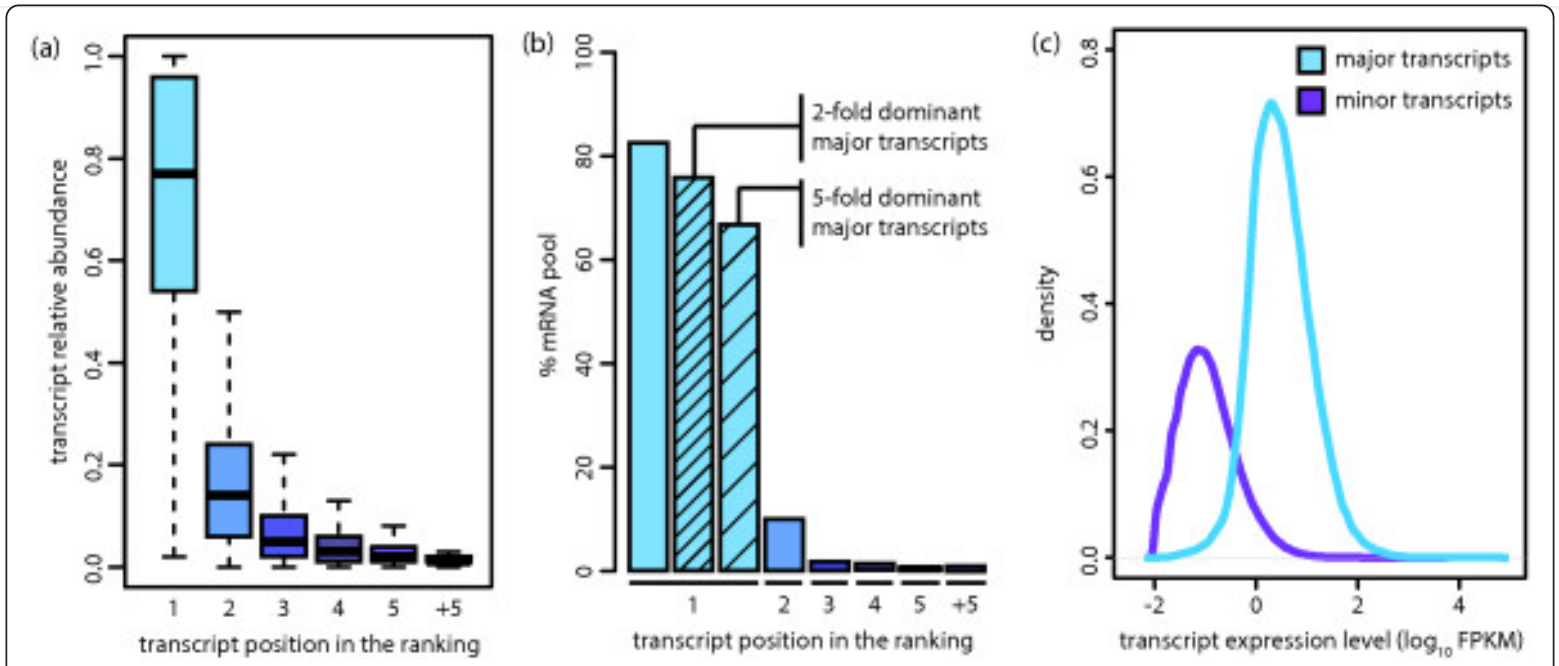
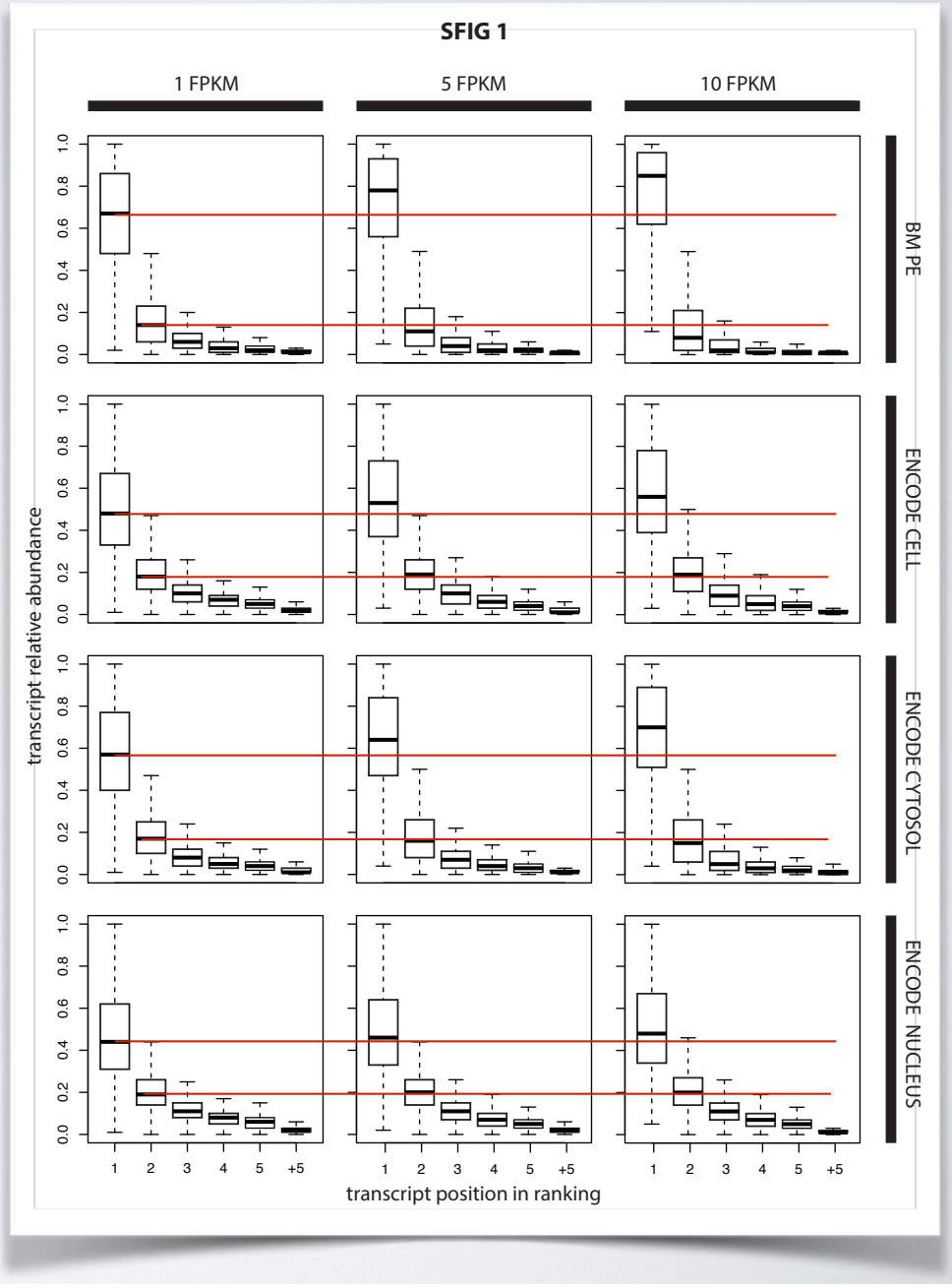


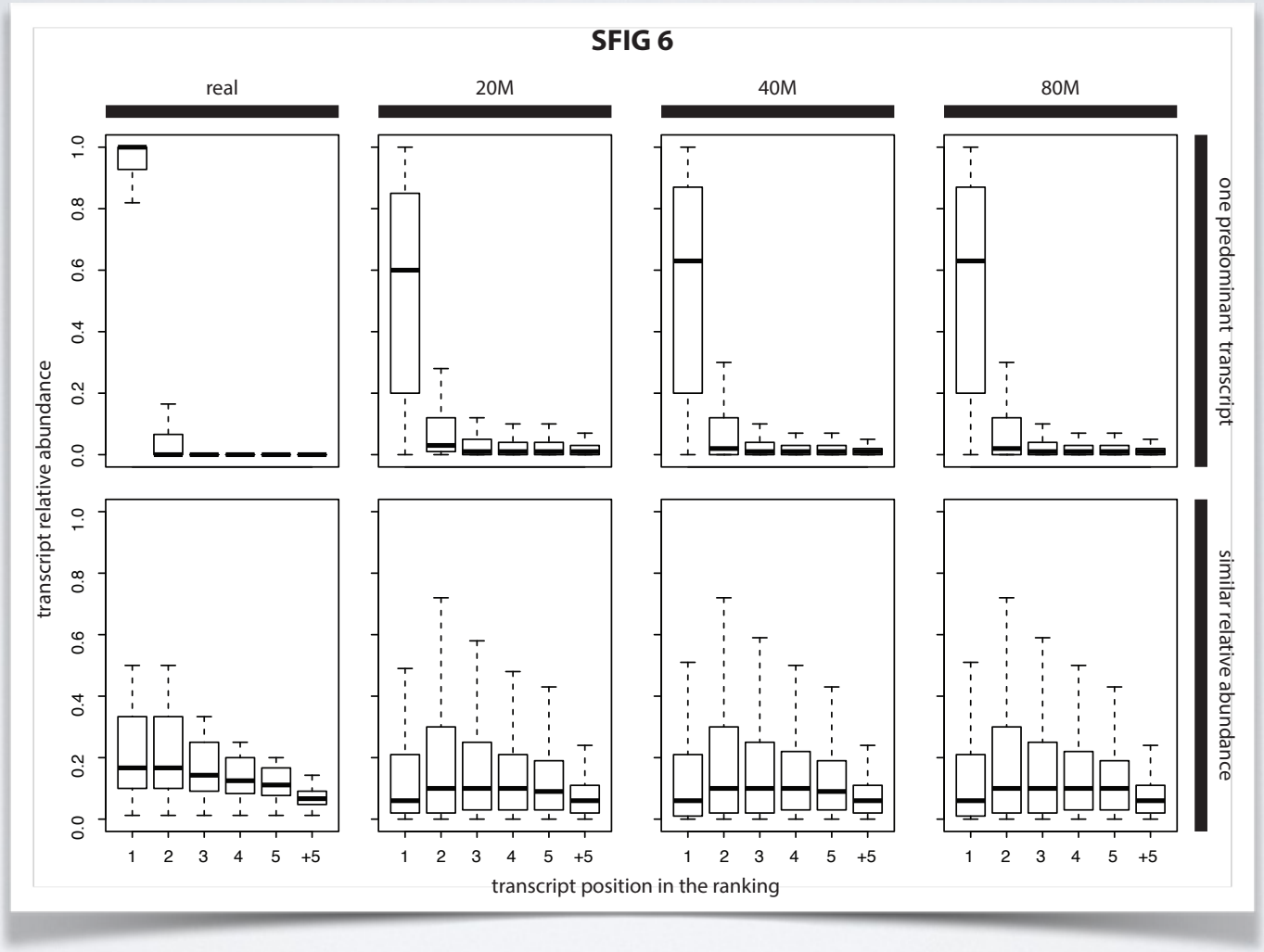
Figure 1 Most protein coding genes express one predominant transcript. (***a***) Relative abundance of the subset of transcripts in each position of the ranking for the primary tissues dataset. For each gene, transcripts were ranked based on their relative abundances. There is generally one predominant transcript over the rest. (**b**) *Percentage of the studied mRNA pool explained by each category of transcripts for the BM dataset. The mean percentage for all samples is represented here. Major transcripts represent approx*imately *85% of the studied mRNA population and were further classified into* two-*fold and* five-*fold dominant*. (**c**) Expression distribution for major and minor transcripts in the tissue dataset. We detect a total of 31,902 transcripts expressed above 1 FPKM in at least one tissue and 26,641 different major transcripts.

body map vs. ENCODE

- major transcript abundance **greater in cytosol** than in nucleus
- major transcript abundance is generally **lower in cell-lines** than in tissues



simulated data



- major transcript abundance is **underestimated** in simulated data

transcript switching

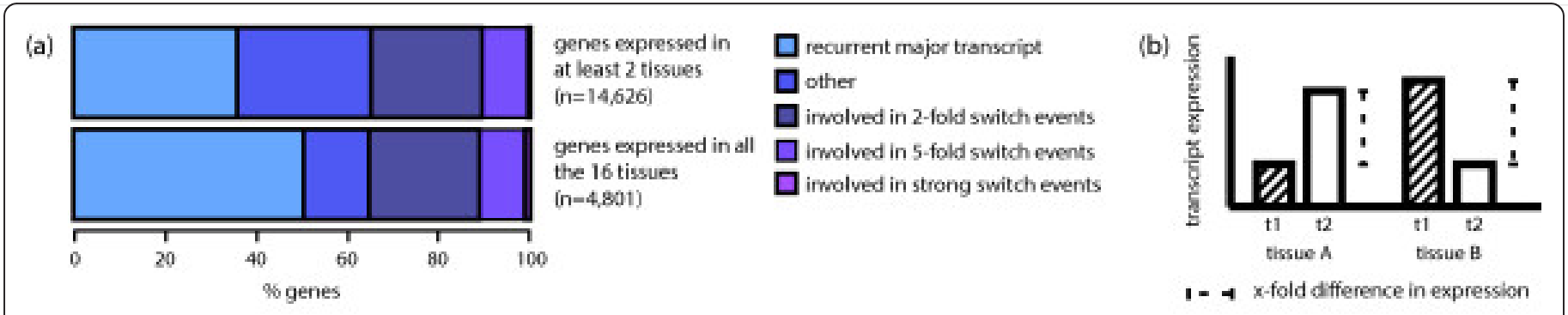
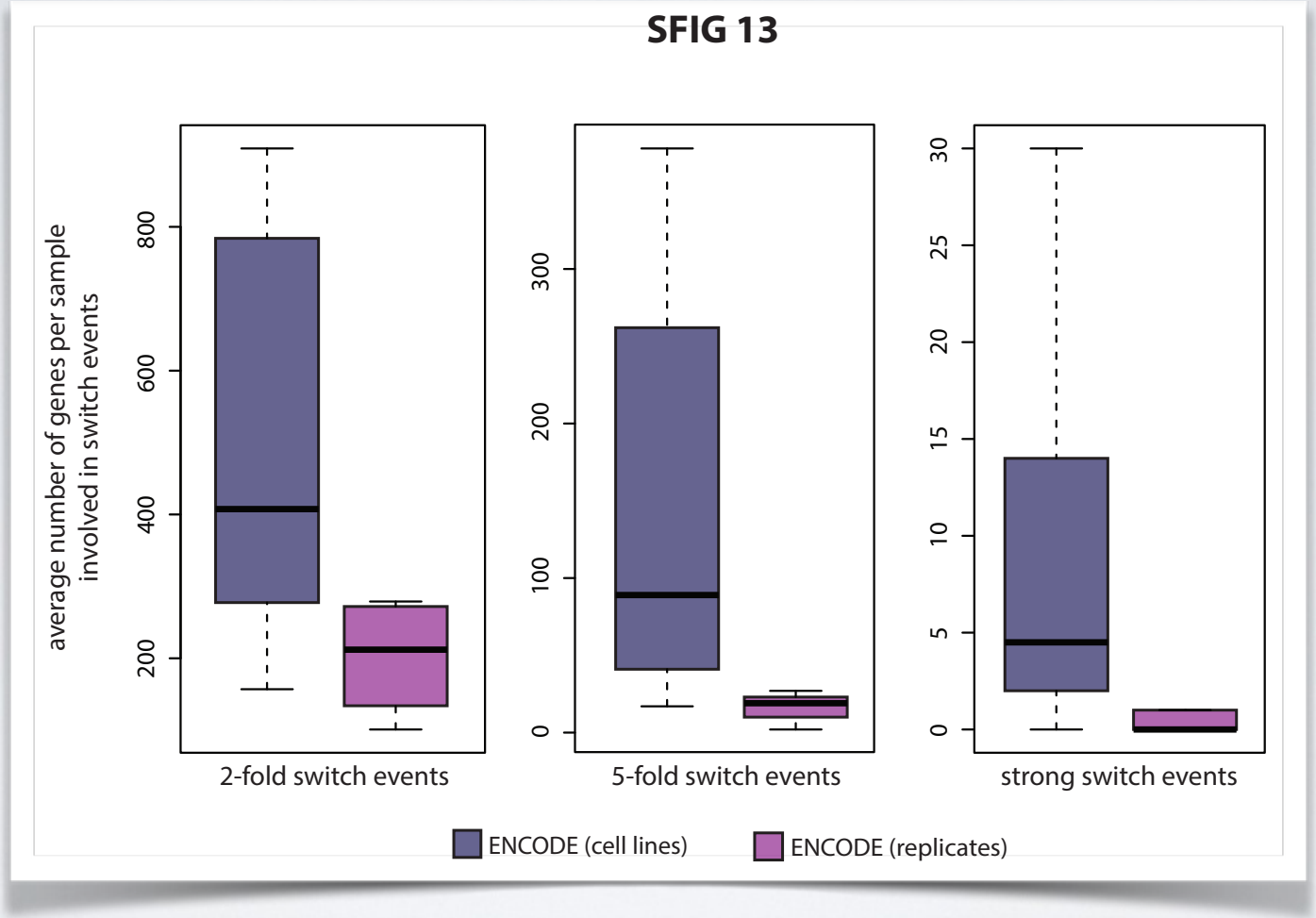


Figure 3 Expression patterns for major transcripts. *(**a**) Percentage of genes with recurrent and non-recurrent major transcripts. Changes in the identity of major transcripts across samples were quantified with switch events*. (**b**) Concept of switch event. A gene is considered to be involved in a switch event if we detect two different dominant major transcripts in two different samples. If the dominant transcripts involved in the switch are expressed above 5 FPKM, while the minor ones are expressed below 1 FPKM, we define the event as a strong switch.

- 50% of genes expressed in all 16 tissues have the same dominant transcript
- 35% of genes have consistent gene-level expression, but have different dominant transcript between at least 2 tissues

replicates



- biological variability greater than between technical replicates

example transcript switch

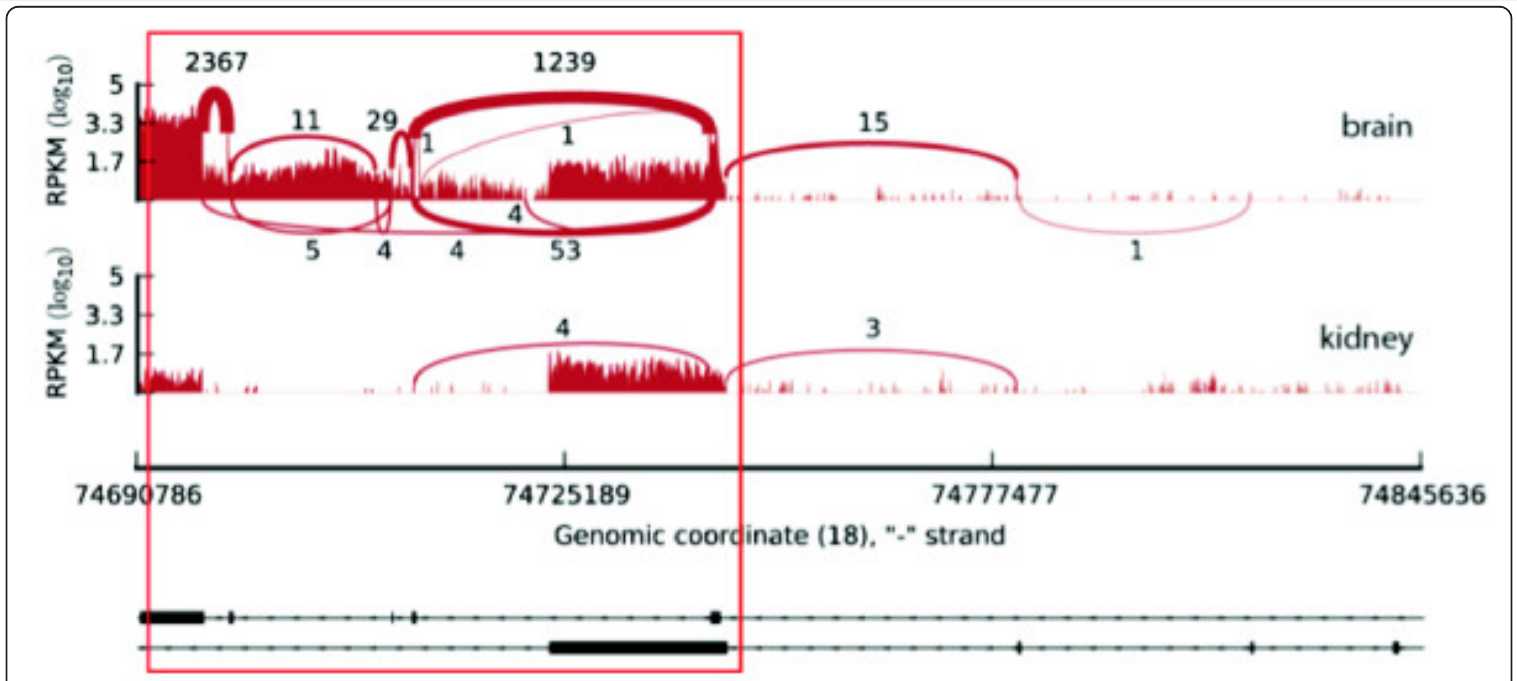
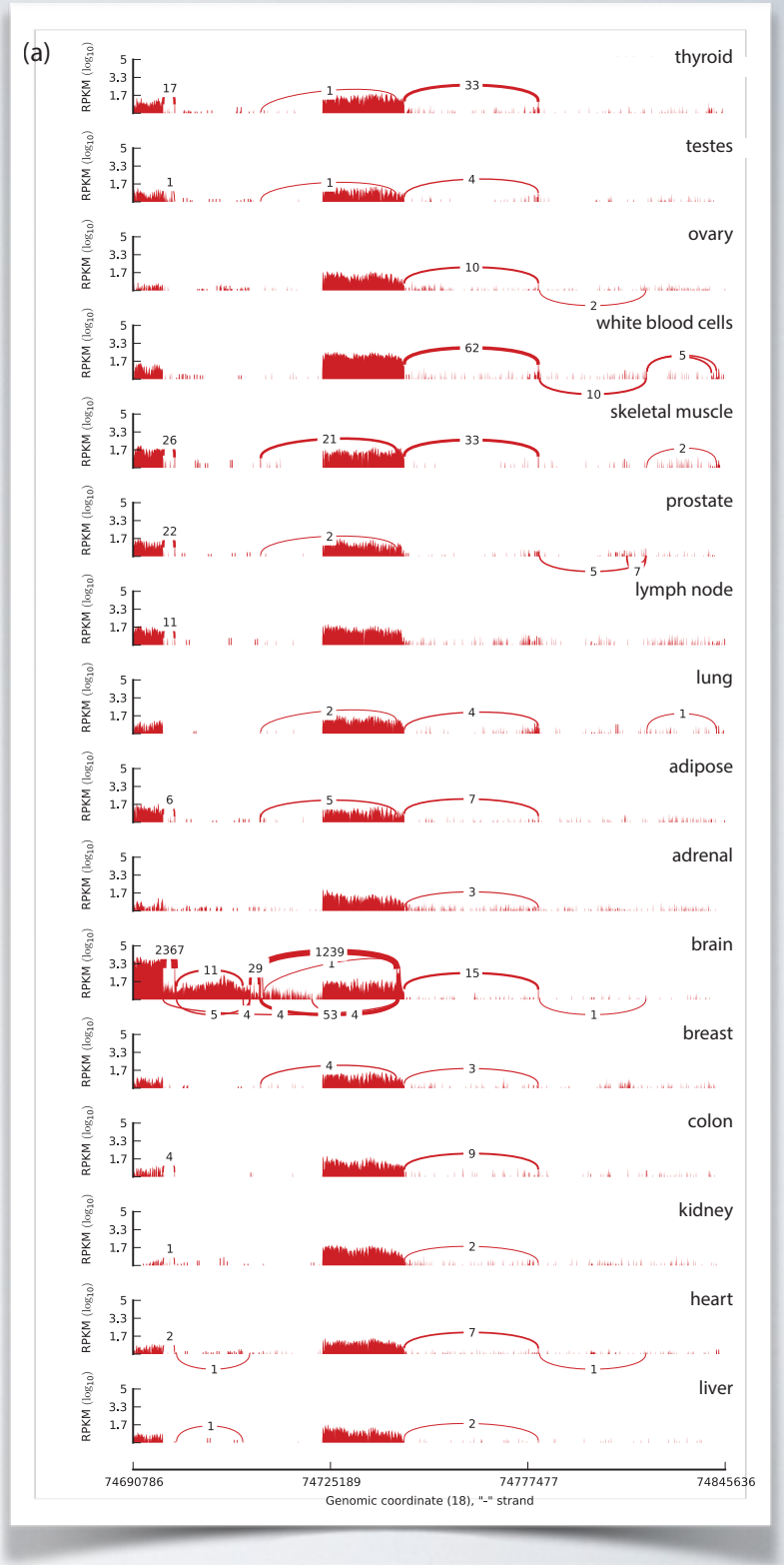


Figure 4 Example of a switch event: *MBP* (myelin basic protein, ENSG00000197971). Read coverage for the gene in brain and kidney. Further tissues, as well as transcript annotation information, can be visualised in Additional File 1 - Figure S12.

example transcript switch

- Myelin Basic Protein (MBP)
- very different transcript in brain compared to all other tissues
- codes for a different protein
- the two proteins almost certainly have different functions (myelin acts as electrical insulation to neurons)



non-coding major transcripts

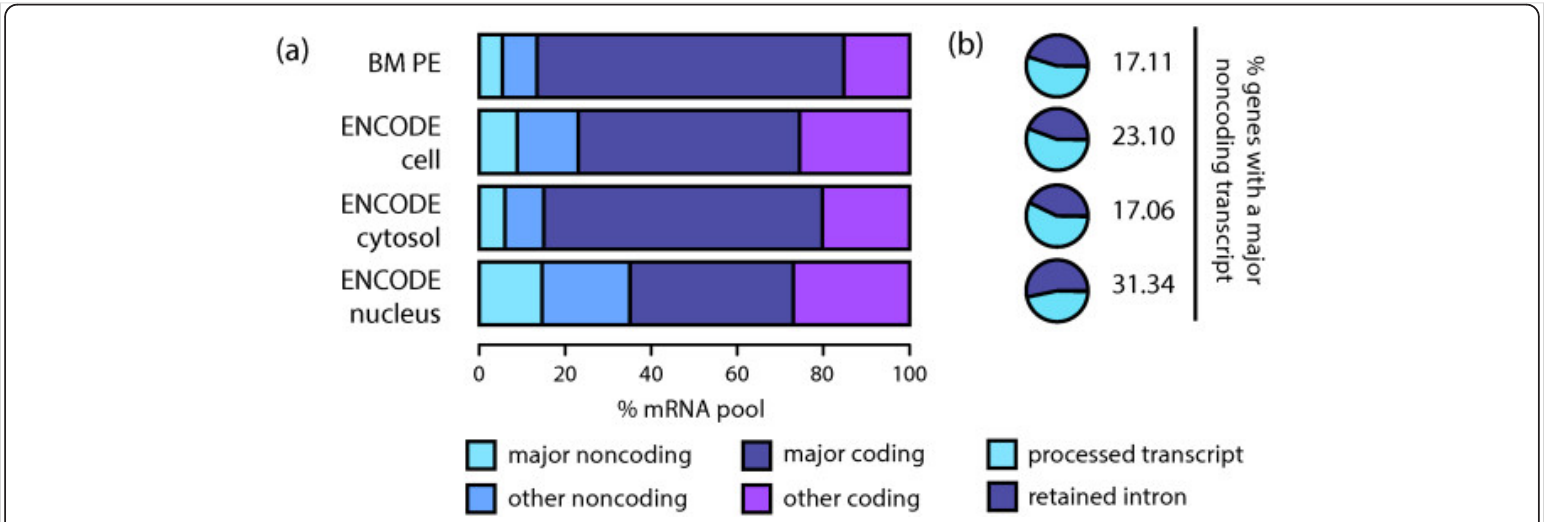


Figure 5 Major non-coding transcripts in protein coding genes. (**a**) *Proportion of the mRNA studied represented by different categories of transcripts. Average proportions were calculated including all the samples from each dataset. Major non-coding transcripts are more abundant in nucleus, where the proportion of major coding ones also becomes reduced.* (**b**) Transcript biotype categories for the major non-coding transcripts. Average proportions were calculated including all the samples from each dataset. Processed transcripts are more abundant in the cytosol, while retained introns represent the major fraction in the nucleus. Other minor categories that represented <1% of the transcripts were also identified, but are not visible in the plots.

- 17% of 'protein-coding' genes have a major transcript that is non-coding (31% in the nucleus)

a JC of two (or three) halves

1. where's miRNA?
   - recap of miRNA/mRNA binding from AGO IP assays

2. one transcript to rule them all
   - transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene

3. neurogenesis in the adult human brain
   - if there's time…

neurogenesis in the adult human brain

image: Fabian Oefner

increased atmospheric ^{14}C

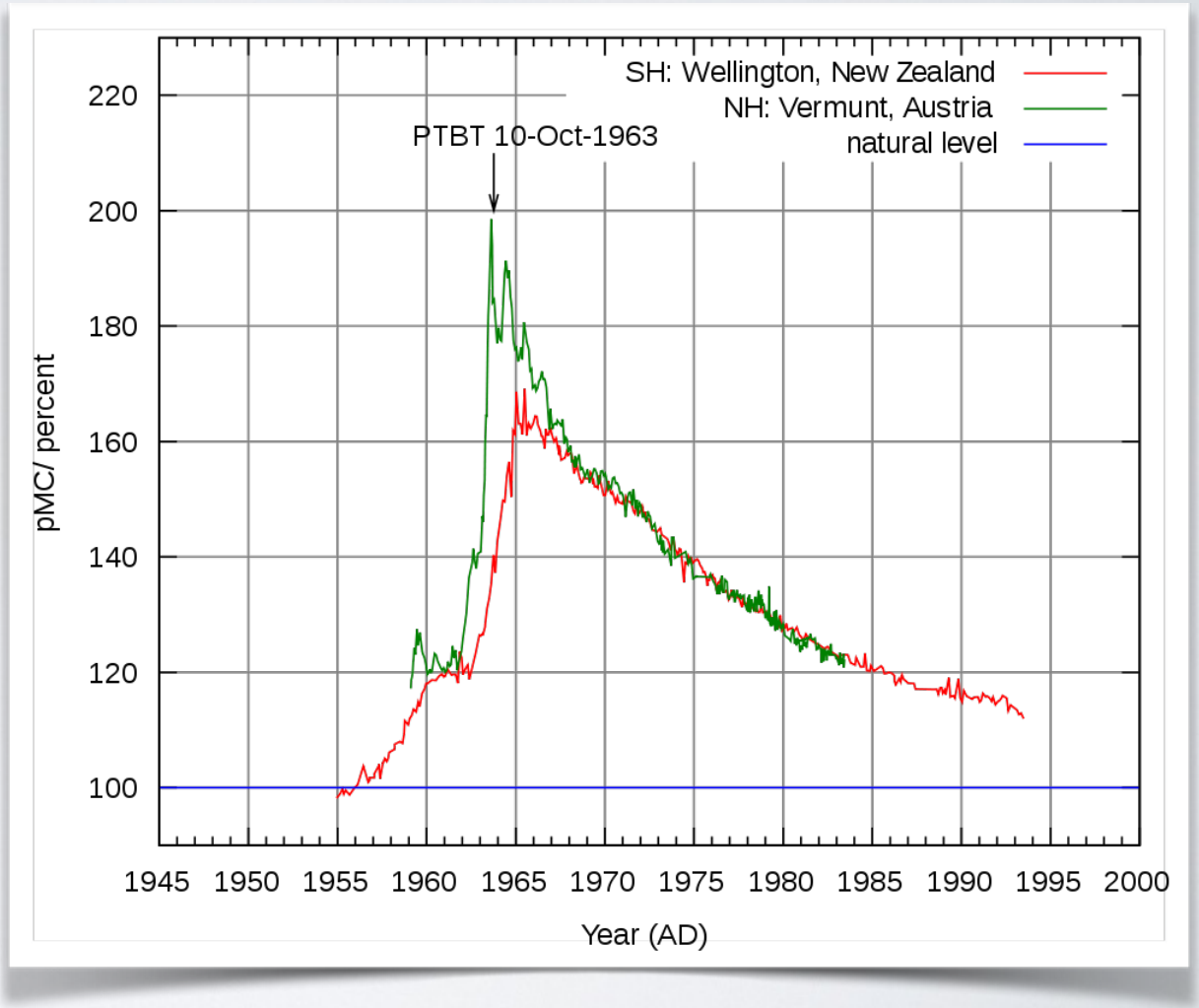
enters food chain

incorporated into tissues

enters entire food chain

radiocarbon (^{14}C) dating

- radiocarbon dating invented in 1949
- principle that atmospheric ^{14}C is constant and ^{14}C decays predictably over ~14,000 years
- nuclear testing started in 1955…



genesis of non-neurons

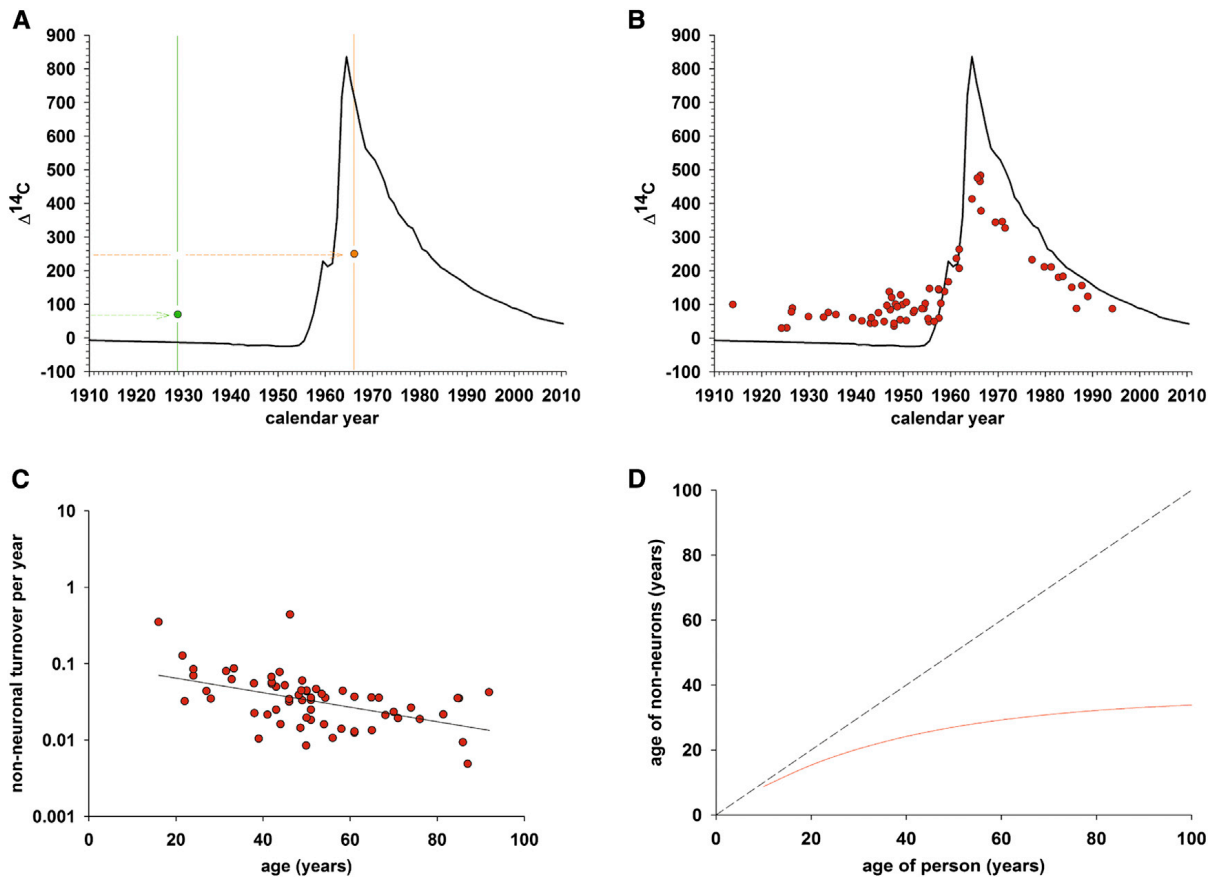


Figure 2. Turnover Dynamics of Nonneuronal Cells

(A) Schematic illustration of the representation of the measured ^{14}C concentration in genomic DNA. The black line indicates the ^{14}C concentration in the atmosphere at different time points in the last century. Individually measured ^{14}C concentrations in the genomic DNA of human hippocampal cells are plotted at the time of the subject's birth (vertical lines), before (green dot) or after the ^{14}C bomb spike (orange dot). ^{14}C concentrations above the bomb curve (subjects born before the bomb peak) and data points below the bomb curve (subjects born after the nuclear tests) indicate cellular turnover.

(B) The ^{14}C concentrations of genomic DNA from nonneuronal cells demonstrate postnatal cell turnover in subjects born before and after the bomb spike.

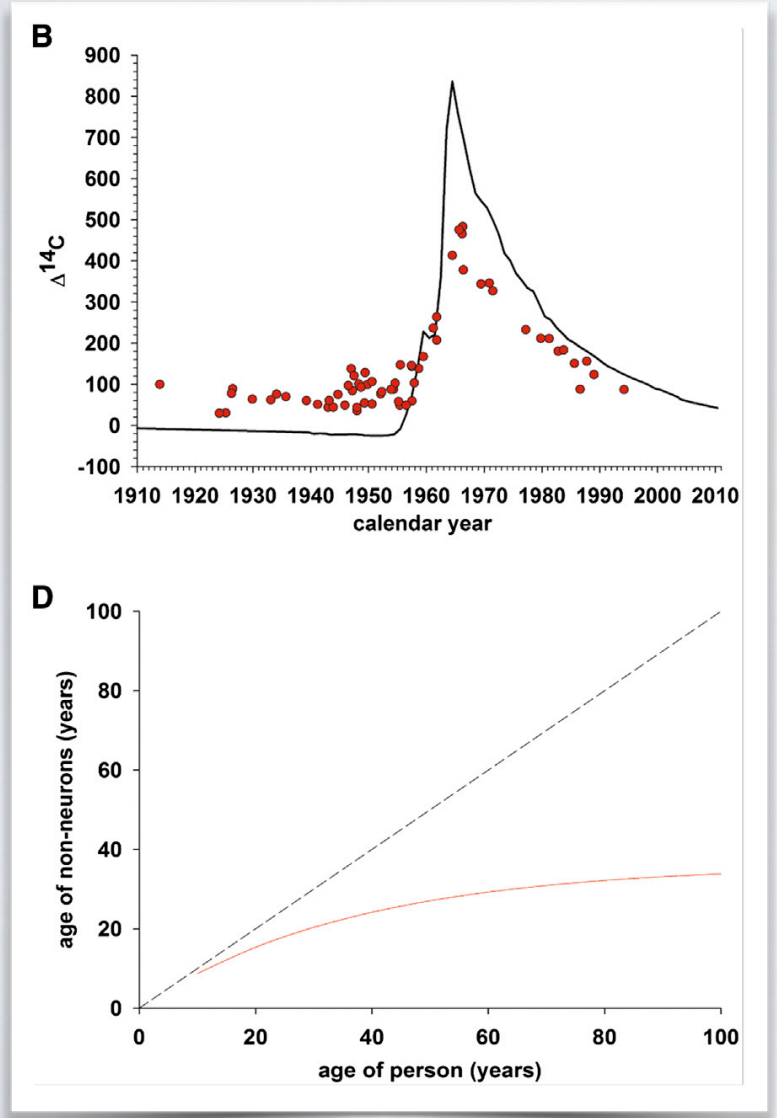
(C) Individual turnover rates for Nonneuronal cells computed on the basis of individual data fitting. Individual turnover rate calculations are sensitive to deviations in measured ^{14}C and values <0.001 or >1.5 were excluded from the plot, but the full data are given in Table S1.

(D) Nonneuronal average cell age estimates of cells within the renewing fraction are depicted (red curve). The dashed line represents a no-cell-turnover scenario.

See also Figure S2 and Table S2.

genesis of neurons

non-neurons



neurons