Endogenous RNA editing records the transcriptional history of human cells.

Ali Essam Ghareeb

University College London

And

Francis Crick Institute

Submitted for the degree of:

Doctor of Philosophy - University College London

1. Declaration

‘I, Ali Ghareeb confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

2. Acknowledgements

BSM.

For Evelyn, Essam, Ashraf and Fouad.

I've been fortunate to have many great mentors, without whom this work would not be possible. If you made my life hard at some point, thank you.

Thank you to the many Crick scientists who offered to lend an ear when I had a challenge I couldn't solve on my own. This work would certainly not be possible without you.

I am grateful for the contributions of James Bayne (PhD student in the Rodriques lab) who developed the code for section 10.4, and Aaron Wagen (PhD student in the Rodriques and Ghandi labs) who developed the code used in section 13.4.

Thanks also to the rest of the Timestamps team – Laura, George, Alya, David and of course Sam.

3. Abstract

Obtaining continuous measurements of transcriptional activity across the transcriptome would unveil a granularity of information about cell regulation not currently possible. RNA sequencing, however, takes a snapshot of the transcriptome at an instant and is destructive, preventing further sampling. From the point of transcription until degradation, mRNA gradually accumulates adenosine-to-inosine edits due to the activity of ADAR enzymes, potentially providing an endogenous record of mRNA age. Determining the age distribution of transcripts present within a cell would allow one to infer the cell's recent transcriptional record.

In this work, I find that endogenous A-to-I editing of RNA transcripts in unmodified human cells serves as a molecular recorder, allowing us to infer the ages of endogenous RNA transcripts at the single molecule level with precision on the order of hours. Using a controlled gene expression system, I show that the distribution of ages of a transcript encode past transcriptional history, information which is not found in static RNA sequencing. I show that Timestamps can recover the transcriptional dynamics of hundreds of genes in differentiating primary human monocytes, even identifying new putative regulators of human monocyte differentiation. In dividing single cells, I show that the periodic transcription of S- and G2/M-phase associated genes, as recorded by Timestamps, tracks the period of the cell cycle.

By sorting the single neurons, Timestamps could enable RNA sequencing to identify which neurones respond to diverse but temporally separated stimuli, thus enabling high-throughput functional brain mapping. I begin developing an *in vivo* experiment to show that Timestamps can trace the production of immediate early genes and thereby infer the past activity of neurons.

Finally, to improve Timestamps and extend its functionality to non-primate model organisms, I invent a new class of promiscuous, hyperactive RNA editors.

Timestamps is an endogenous molecular recorder of transcriptional history.

4. Impact Statement

The ability to continuously measure activity within human cells will illuminate the gaps in our knowledge of how cells achieve fine regulation over innumerate complex processes. Timestamps enables researchers to make continuous measurements of the transcriptomic dynamics in a non-invasive manner. As I have demonstrated in this thesis, this technology differentiates processes which occur with different kinetics (fast versus slow), can tell when genes were first transcribed (e.g. 2 hours ago) and identifies transient yet critical mechanisms of transcriptional regulation which are missed when measurements are taken at a single endpoint.

In the realm of academia, researchers now have an accessible and highly multiplexed method for making measurements of transcriptional dynamics. By being identifying endogenous editing by ADAR enzymes as a rich source of temporal information, I believe it will not be long before other research labs pursue strategies of using other RNA modifications, such as methylation to improve the Timestamps method. Yet more labs will work on developing more hyperactive, promiscuous editors of RNA. Thus, the novel work presented herein has launched a new field of RNA-modification based temporal transcriptomics which I believe will see a spate of new publications.

Because continuous measurements can be taken from a single sample, this will enable experiments which were previously not possible. For example, experiments in which animal tissue is sampled and prepared for sequencing. Timestamps enables scientists to interrogate experiments where multiple temporally separated perturbations are applied to a single animal.

One can envisage using this technology to identify new mechanisms of regulation within human cells which could make promising drug targets. The fact that Timestamps works in a

label-free manner makes this even more accessible in primary cells which are notoriously difficult to engineer.

5. Table of Contents

6. List of Figures

7. List of Tables

8. Introduction

The purpose of this section is to introduce the reader to the foundational work which is vital for understanding the experimental work which follows. Principally, this foundational work is in the fields of RNA editing and Next Generation Sequencing, and the nascent field of Temporal transcriptomics, where this chapter will begin. I will conclude the section with a discussion of human Adenosine Deaminase Acting on RNA (ADAR) enzymes and the prevalence and breadth of their activity.

8.1. Temporal Transcriptomics

The transcriptome is a highly dynamic population of molecules whose composition reflets cells internal state. The determination of the sequence of events in living systems poses a challenge. One possible approach involves the direct observation of interactions using live-cell fluorescence microscopy, which allows for real-time ordering. Another approach entails performing time series experiments by destructively sampling and transcriptionally profiling the system at various time points. Epistatic analysis represents a third approach that involves the ordering of gene actions through the comparison of single and double mutants' phenotypes (Choi et al. 2022). Despite the strengths associated with these and other approaches, they are also subject to significant limitations. For example, the number of simultaneous recordings which can be made by live microscopy is limited by spectral overlap, and in time series experiments, the granularity of temporal information is limited by the frequency of sampling. As sequencing is a destructive process, it provides only a snapshot of the transcriptome at a single point in time. Previous attempts to encode temporal information about the cell's transcriptional state have relied on three encoding formats: DNA, RNA, or protein.

8.1.1. DNA-based recorders

DNA-based recorders rely on a system of signal sensing (for example the production of a DNA writer under an inducible promoter), DNA writing (CRISPR integrases, CRISPR-Cas9, Site-specific Integrases and Base editors), and DNA reading (usually DNA sequencing, but also fluorescent probes) (Sheth and Wang 2018). The CRISPR-Cas system is the prokaryotic adaptive immune system. DNA sequences of pathogens are integrated into the host genome to form a memory which may be used to identify the and destroy the same sequences in the future. Early attempts at temporal encoding cell state information harnessed the Cas1-Cas2 integrase to sequentially integrate synthetic oligonucleotides into CRISPR arrays (Shipman et al. 2016). Oligonucleotides consisting of a PAM and 32 variable nucleotides were electroporated into E.coli. The order of electroporation was recorded in the E.coli genome and subsequently retrieved by DNA sequencing. In a similar vein, recombinases can be used to perform irreversible operations on a plasmid-based DNA register in response to chemically-induced stimuli (Roquet et al. 2016). In theory, recombinases could be co-regulated under an endogenous promoter to record gene activation. However, both CRISPR and site-specific recombinase approaches require significant engineering to implement in mammalian systems. For example, mammalian cells would need to express the Cas enzymes, and any reporter oligonucleotides would need to be engineered with the appropriate PAM sequence to be recognised and integrated into the host genome. Although these methods produce an indelible and easily interpretable readout, they are difficult to implement in mammalian systems and difficult to scale hundreds of genes simultaneously. Furthermore, they produce only binary recordings of cellular events, not encoding for example, transcriptional kinetics. More recent improvements have expanded the information encoding capacity of DNA writers (Loveless et al. 2021; Choi et al. 2022), and enabled recording of the length of transcriptional events (Loveless et al. 2021) however the fundamental problems of multiplexed gene monitoring and implementation in mammals remains.

Overall, DNA-based recording devises are unlikely to be useful in recording cellular events across multiple genes and in vivo, however they may be more suited to cell lineage tracing (Kalhor, Mali, and Church 2017; Loveless et al. 2021).

8.1.2. Protein-based recorders

Records of cellular events may also be encoded in self-assembling protein fibers as fluorescent tags (Linghu et al. 2023; Lin et al. 2023). Lin et al. use a fusion of the Pak4 kinase and a 38 amino acid domain of its inhibitor. This protein, iPAK4, forms a growing protein fibre which is decorated with two types of fluorescent timestamps. The first, a HaloTag-iPAK4 fusion protein, is incorporated into the growing fibre at regular intervals and acts as a fiduciary timestamp marking the passage of clock time. The second, eGFP-iPAK4 is under the control of a promoter of interest (e.g. cFOS) and acts as a reporter of intracellular activity. Halotag fusion proteins form covalent bonds with Halotag ligands, haloalkane-dyes which cannot be genetically encoded (Los et al. 2008). Expressing this system in mouse primary neurons, the authors monitored the transcriptional activity of the Immediate Early Gene (IEG) cFos after stimulation with PMA. The authors observed green fluorescent bands in the fibres corresponding to two epochs of PMA stimulation and where able to take measurements for up to 22 hours with a time resolution of 30 minutes which is limited by the spacing of the fiducial timestamps (itself limited by the 4.5 hour half-life of the HT-iPAK4 fusion protein (Lin et al. 2023).

Linghu and colleagues have developed a similar system for tracking transcriptional activity in cells called "expression recording islands" or XRI. Using self-assembling protein chains and fluorescence imaging, the XRI system is like the ticker-tape system previously described by Lin and colleagues. However, the XRI system is fully genetically encoded, meaning no dyes are necessary for its operation.

The researchers harnessed the power of a mutant form of a protein called 1POK(E239Y), which is an engineered filament-forming protein inspired by an E. coli isoaspartyl dipeptidase. Through the fusion of the 1POK(E239Y) protein with an "insulator" component such as mEGFP or maltose binding protein (MBP tag), the authors ensured that protein assembly resulted in fiber-like structures instead of protein aggregates. Both methods are capable of single-cell measurements, with time resolutions on the order of tens of minutes and total recording times on the order of days. Multiplexed measurement of several genes is limited by the spectral overlap of reported fluorochromes, and possibly by competition between competing subunits. Furthermore, readout of the spectral record by fluorescence confocal microscopy is low throughput.

8.1.3. RNA-based methods

Metabolic pulse-chase methods such as SLAMseq (thiol-linked alkylation for the metabolic sequencing of RNA) have attempted to introduce a temporal dimension to RNAseq data by introducing metabolic labels which are incorporated into nascent mRNA molecules during transcription (Herzog et al. 2017; Muhar et al. 2018). Cells are grown in the presence of a thiol-containing nucleotide analogue, such as 4-thiouridine. This analogue is rapidly incorporated into nascent RNA during transcription. Total RNA is isolated and incubated with iodoacetamide which covalently attaches a carboxyamidomethyl group to s4U by nucleophilic substitution, labelling almost all 4-thiouracil nucleotides. Thiol alkylation results in T>C conversions during reverse transcription, thereby providing a readout of thiol incorporation in sequencing data. Unlike the DNA and protein recording methods described above, SLAMseq allows parallel observation the entire transcriptome and can measure global and transcript-specific transcription rates and mRNA half-lives. However, this method only provides temporal information relative to a single point in time (the time at which the 4-thiouridine was added) and the temporal resolution is low (hours) owing to the slow absorption and incorporation of 4-thiouridine. Although, in vivo implementation of SLAMseq is possible in zebrafish (Bhat et al. 2023), implementations in murine models have focussed

on labelling cell types of interest (to supplant the need for FACS sorting) rather than temporal transcriptomics (Matsushima et al. 2018).

The observation that 15-25% of reads from RNAseq data generated through a poly-T primed reverse transcription step (thus selecting for mRNA) contain unspliced intronic sequences led to the development of RNA velocity (La Manno et al. 2018). RNA velocity is a computational method used to infer the direction and rate of gene expression changes in single cells. It can be used to understand the dynamics of cell state transitions, such as cell differentiation, during development or disease progression. RNA velocity employs a simple kinetic model of transcription, splicing and degradation. The authors applied this model to a published compendium of mouse tissue RNAseq data and found that most genes across a wide range of cell types can be modelled at steady state with a similar rate of degradation. By assuming this fixed rate of degradation, the ratio of unspliced to spliced transcripts can inform weather a gene is being induced or repressed at that point in time.
Of note, while the DNA and protein recorder methods described above record the occurrence of cellular events on a single ledger, the RNA-based methods record the ages of single RNA molecules. Thus, in RNA-based ledgers, the recording is distributed across the transcriptome, and the record of events is the distribution of ages of the mRNA molecules. Therefore, methods which capture only the transcript ages relative to one timepoint (SLAMseq) do not capture the entire history of a single cell.

8.1.4. RNA Timestamps

RNA Timestamps overcomes the limitations of SLAM-seq and RNA velocity by encoding the ages of single mRNA molecules as edits which were introduced by a hyperactive ADAR mutant expressed in trans. By counting the numbers of edits on molecules of RNA using RNA-sequencing (see section 8.3), the authors were able to infer the ages of single molecules of RNA (identified by unique molecular identifiers). Of course, inferring the ages of molecules of RNA is by itself, unlikely to be useful. Next, the authors showed that by

obtaining the ages of single molecules of RNA (produced under the action of a single promoter), they were able to build up a picture of the transcriptional history of the cell from that promoter. This is analogous to building up a picture of the paleontological record from carbon dating of fossils.

Editing construct

The authors screened three ADAR variants in order find a pair with an editing timescale on the order of hours. They screened 3 mutants of ADAR2: E488Q, T490A and E488Q/T490A. The E488Q and T490A mutations are in a conserved loop within the active site of ADAR2. The E488Q mutant has been shown to have relaxed sequence preference and increased catalytic rate (promiscuity and hyperactivity) (Kuttan and Bass 2012). Meanwhile, the T490A mutant did not show reduced sequence preference, and in fact had a reduced rate of catalytic activity in previous studies so it is not clear why the authors chose to study this mutant. In fact, the same previous work concluded that T490 is critical for an essential base-flipping step in ADAR2 catalysis. Additionally, the T490A mutant abrogates the hyperactivity conferred by E488Q, so it is not clear why the E488Q/T490A double mutant was studied if the aim of the authors was to identify an optimal editing rate (Cox et al. 2017; Kuttan and Bass 2012).

Recording edits

A ledger was developed to record adenosine-to-inosine (A-to-I) edits. This RNA ledger (referred to as a Timestamp), reported its age via the "gradual accumulation of A-to-I edits caused by an engineered version of the human adenosine deaminase acting on RNA 2 catalytic domain" (Rodriques et al. 2021). The ledger was designed as arrays of adenosines arranged to be an ideal substrate for the mutant ADAR2 catalytic domain. Edits in the ledger are identified as A-to-G mutations in next generation sequencing data.

The mutant ADAR2 catalytic domain is recruited to the RNA ledger by MS2 binding sites placed adjacent to the ledger (Figure 1A and B). The MS2 protein, from the MS2

bacteriophage, binds to a stem-loop structure present in the MS2 single-stranded RNA genome. This phenomenon is exploited in MS2 tagging, originally developed to track the movement of RNA molecules in yeast over time (Bertrand et al. 1998).

Editing Model

To calibrate a model of numbers of edits (on the ledger) versus time, the authors devised a scheme where a TET-ON promoter was induced using doxycycline and 1 hour later transcription was halted using actinomycin D. The cells are lysed at regular intervals thereafter (Figure 1C). This has the effect of producing a population of mRNAs with a known mean age. A model describing the numbers of edits on the ledger as a function of time can then be parametrized from this experiment. For each editing site on the RNA ledger, the authors fitted an exponential cumulative distribution function (CDF) (Equation 1) which described the accumulation of edits with time. Retaining only sites with an R2 value of 0.9 (roughly half of all editing sites), it was found that the distribution of edits on the ledger was well-described by a Poisson binomial model (Equation 2), where the number of trials is the number of editable positions on the ledger and success probabilities at each position are given by Equation 1. Using this approach, the authors were able to infer the age of single mRNA molecules on the scale of hours from only the empirical distributions of RNA edits per molecule.

However, this approach is difficult to more than a few mRNAs due to the requirement for engineering an RNA ledger into the target RNA and the need for an MCP-ADAR2 editing construct. Furthermore, *in vivo* implementations would require genetically engineered organisms expressing the MCP binding motif in mRNAs of interest.

In summary, Rodriques et al. have shown that the ages of individual RNA molecules can be recorded with hour-scale fidelity by the activity of an engineered adenosine deaminase on an engineered RNA substrate. In principle, this method can be used to infer transcriptional dynamics non-invasively with hour-scale time resolution from sequencing data collected at

an endpoint. However, this method is limited by the need for both an engineered deaminase and an engineered RNA substrate.

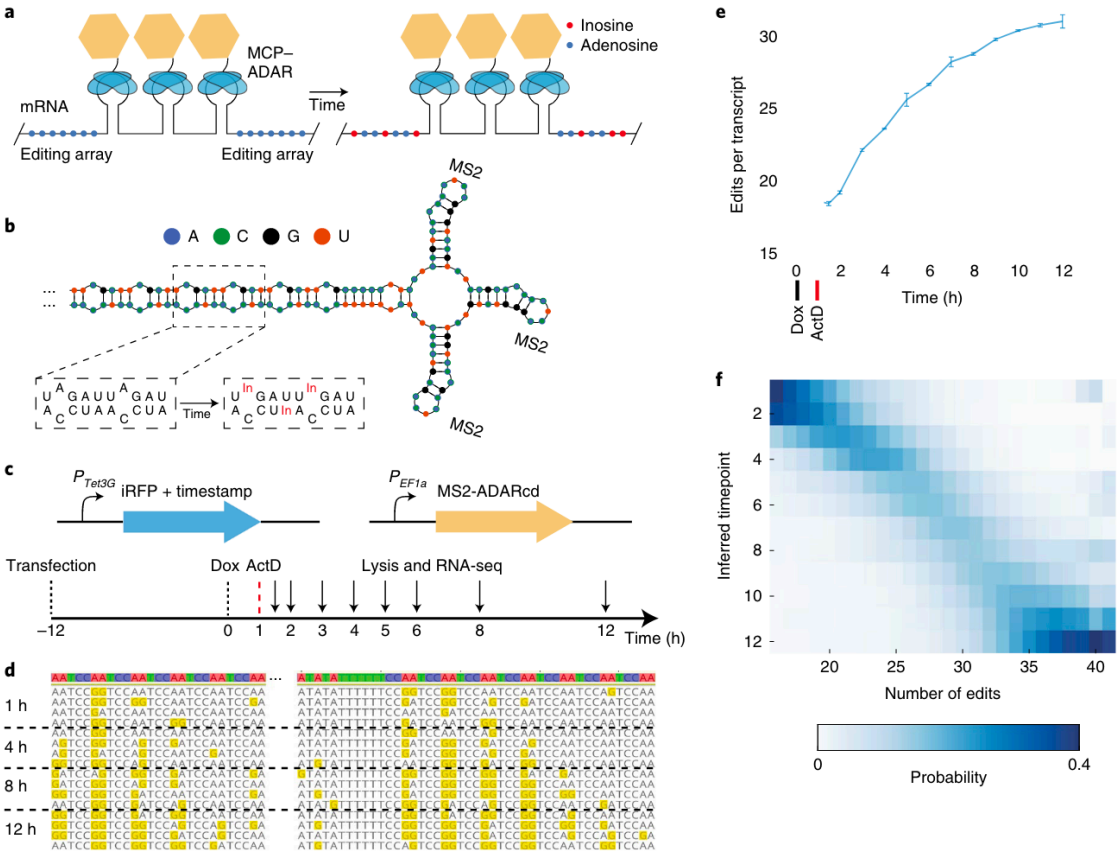


Figure 1 Timestamps 2021 – RNA edits record time (reprinted with permission; (Rodriques et al. 2021))

(a) MCP-ADAR is recruited to the RNA ledger via loop-stem hairpins and edits in the adenosine-rich regions of the ledger.

(b) MCP-ADAR preferentially edits at A::C mismatches in the RNA ledger.

(c) A timestamped red fluorescent protein and the MCP-ADAR editing construct were expressed in trans. A timeline of the experiment is shown.

(d) Examples of RNA ledgers edited for different periods. Older ledgers have more edits.

(e) Mean number of edits versus time from doxycycline induction.

(f) Heatmap showing the relationship between numbers of edits on a ledger and the probability of being a given age.

8.2. RNA editing

RNA editing involves discrete changes to the nucleotide sequence of an RNA after transcription has taken place. Because editing takes place in RNA, and not DNA, RNA editing allows flexible recoding of codons, isoform switching, premature termination and affects RNA secondary structure. RNE editing, found across all kingdoms of life, was first discovered in mitochondrial mRNA of trypanosomes (Benne et al. 1986). RNA editing is prevalent in the mitochondria and chloroplasts of plants. For example, in flowering plants, mitochondria exhibit extensive RNA editing, primarily involving the conversion of cytidine (C) to uridine (U) (Covello and Gray 1989).

The most common form editing observed in animals, however, is A-to-I editing which takes place due to the expression of a highly conserved class of enzymes which are unique to animals; Adenosine Deaminase Acting on RNA (ADARs). C-to-U editing by deamination also takes place, but to a lesser extent (citation) and is catalysed by cytidine deaminases.

RNA editing appears to play different roles across Species, and accordingly has different transcriptomic distributions. In mice, A-to-I edits are found mostly in introns (approximately 80%) (Licht et al. 2019). In cephalopods and *Drosophila*, A-to-I editing is more commonly found in coding regions, and more likely to be non-synonymous than expected by chance (Yablonovitch et al. 2017; Alon et al. 2015; Graveley et al. 2011). In the octopus, A-to-I RNA editing provides a rapid means for inducing neural adaptation to temperature change. Upon a cold temperature shift, the *Octopus bimaculoides* nervous system induces tens of thousands of A-to-I edits resulting in ~13,000 codon substitutions (Birk et al. 2023).

In primates and humans, the distribution of A-to-I editing is unique due to the presence of dsRNA elements called Alu repeats (Tan et al. 2017; Mansi et al. 2021; Bazak et al. 2014). Almost all (99%) of A-to-I editing sites are found within the introns and 3' Untranslated region (3' UTR) of mRNAs, within Alu repeats, which due to their predilection for forming dsRNA structures, provide a suitable target for ADAR editing (Bazak et al. 2014). Few editing sites in

humans are conserved in other mammalian species. The sites that are conserved tend to be located in genes encoding synaptic proteins or neurotransmitter receptors and are highly edited relative to most sites (Pinto, Cohen, and Levanon 2014). This fact underlies a common theme of editing supporting plasticity in the nervous system; a phenomenon also seen in cephalopods (Yablonovitch et al. 2017).

The most well understood, and first discovered A-to-I editing site in humans is the editing of the GluR2 subunit mRNA which is edited at position 607 introducing a glutamate-to-arginine substitution and thereby reducing the AMPA receptor's permeability to calcium ions (Jonas and Burnashev 1995; Hume, Dingledine, and Heinemann 1991). However, while the functions of a handful of editing sites are understood, the function of most editing sites in humans is unknown and they mostly appear to have no functional significance (Pinto, Cohen, and Levanon 2014; Yablonovitch et al. 2017; Schaffer and Levanon 2021).

8.2.1. RNA editing in humans and non-human primates.

ADAR enzymes

Adenosine Deaminases Acting on RNA (ADARs) are the family of enzymes which deaminate Adenosine in RNA to Inosine (Figure 2) and are found in all metazoans. They were discovered serendipitously by Brenda Bass during her postdoc while trying to understand why antisense RNA knockdown failed to repress gene expression in the embryos of *Xenopus laevis,* the transcriptome of which shows high levels of RNA A-to-I editing (Brenda L. Bass 2024; B. L. Bass and Weintraub 1988). They are expressed in both the nucleus and cytoplasm (PubChem, n.d.).

In humans and non-human primates, three isoforms of ADAR exist: ADAR1 and ADAR2 (also known as ADARB1) are the two functional orthologs, while ADAR3 does not have editing activity and may be an inhibitor of editing by competing for dsRNA binding sites (Tan et al. 2017; Figure 2). Targets for ADAR1 editing are mostly in non-coding, repetitive regions, while

targets for ADAR2 are mostly found within coding sites, however the extent of target overlap is poorly understood and may be substantial (Tan et al. 2017). ADAR1 has two isoforms, p110 (110 kDa) and p150 (150 kDa) (Figure 2). The ADAR isoforms are produced by an alternative splicing event which selects either an upstream (p150) or downstream (p110) start codon (Liu et al. 1997). Additionally, the p110 isoform is co-expressed with p150 from the canonical p150 mRNA due to leaky ribosome scanning. A strong Kozak consensus downstream of the p150 start codon (upstream of the p110 start codon) suggests that the p150/p110 mRNA is designed to leak p110 independently of the expression of p150 (Tony Sun et al. 2021). This mechanism ensures that while the p150 isoform is produced in an interferon-responsive manner, the p110 isoform is constitutively expressed (Patterson and Samuel 1995). Introducing mutations between the p150 and p110 start codons which abrogate p110 expression allows one to study the respective contributions of thep150 isoform is isolation. These studies suggest that half of editing sites are edited exclusively by p150 while the other half are edited by either isoform (Tony Sun et al. 2021).

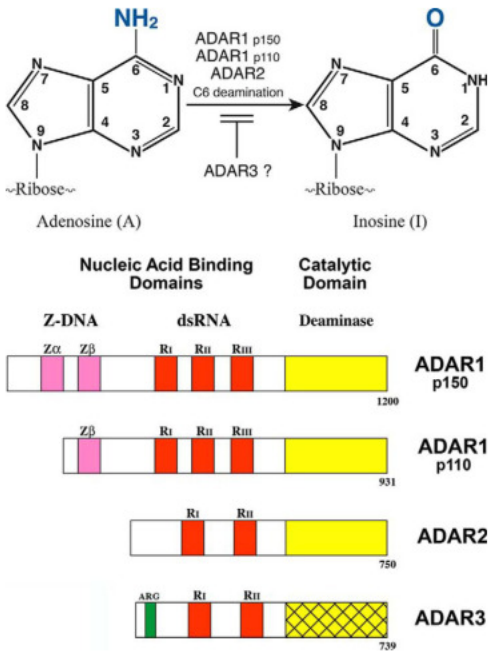


Figure 2 ADAR has three active isoforms and one inactive isoform (shared under Creative Commons CC-BY license; (Samuel 2019)).

ADAR1 p150 modulates the interferon (IFN) response by editing RNAs that would otherwise trigger IFN production through sensors like RIG-I and MDA5 (Pestal et al. 2015). This regulation is vital in balancing immune responses during viral infections and maintaining self-tolerance (Samuel 2019). It is cytoplasmic editing activity by p150 that is critical for inhibiting the immune response to dsRNA. Engineering p110 or ADAR2 to localise to the cytoplasm can substitute for p150 in this respect (Tao Sun et al. 2022).

ADARs are comprised of an C-terminal catalytic domain and multiple N-terminal double-stranded RNA binding domains (Kuttan and Bass 2012). The highly conserved catalytic domain of ADAR is capable of deaminating target adenosines *in vitro* without any co-factors, although target adenosines must be within a duplexed region. The deamination of target adenosines occurs through a base-flipping mechanism which allow the ADAR catalytic domain to access adenosines within dsRNA (Matthews et al. 2016).

Target adenosines occur in the context of preferred 5’ and 3’ nearest neighbours. These preferences are similar for ADAR1 and ADAR2, and are apparent *in vitro* with only purified enzyme and substrate, as they are almost entirely a result of the catalytic domains preferences (Wang, Park, and Beal 2018; Polson and Bass 1994). The most striking preference is an avoidance of adenosines with a 5’ guanosine neighbour (Polson and Bass 1994). Wang et al. 2018 found that the 5' RNA binding loop within the deaminase domain plays a crucial role in determining the target sequence specificity of ADAR1 and ADAR2. They showed that swapping the 5' RNA binding loop from ADAR2 onto ADAR1 resulted in the ADAR1 deaminase domain acquiring the substrate specificity of ADAR2. This indicates that this RNA binding loop is a key determinant of sequence specificity, while the dsRNA binding domains contribute to overall substrate binding affinity their influence on target sequence specificity is small (Wang, Park, and Beal 2018).

The Genotype-Tissue Expression (GTEx) project is a comprehensive research initiative aimed at understanding how genetic variation affects gene expression and regulation across different human tissues ("A More Personal View of Human-Gene Regulation" 2017). Landmark work by Tan and co-authors characterised the tissue-specific distribution of ADAR activity in humans using the most comprehensive list of editing sites in the human transcriptome to date (Tan et al. 2017).

To study the tissue-specific distribution of editing, >8000 RNA-seq datasets from human tissues derived from the GTEx were used. Previous studies, limited by sequencing depth (see section *'detecting editing sites'*, below), tended to identify only the most highly edited sites, which tended to be those found in coding regions in genes with neuronal functions (Pinto, Cohen, and Levanon 2014). Thus, it was thought that most RNA editing took place in neurons in the central nervous system (Ramaswami et al. 2013; Paz-Yaacov et al. 2010). However, most sites are edited at a low level (<1% of reads) and therefore are unlikely to be detected in shallow sequencing (low depth coverage of the target transcriptome means that rare edits are likely to be missed) (Pinto, Cohen, and Levanon 2014; Bazak et al. 2014). Tan et al. pooled the sequencing results of many human samples from their own database and from GTEx. They validated new sites using microfluidics-based multiplex PCR and deep sequencing. Contrary to the belief that most RNA editing activity is found in the brain, analysis of the GTEx data demonstrated broadly similar levels of editing in repetitive sites (non-coding) and coding sites throughout the body (although there was more differentiation amongst tissues in coding sites) (Tan et al. 2017). The exceptions were skeletal muscle which showed distinctly lower levels of editing at non-coding sites, and arteries, which showed distinctly higher levels of editing in coding sites (higher even than the brain).

Many viral RNA genomes contain double stranded RNA (dsRNA) structures, and hence eukaryotic cells have developed innate immune sensing mechanism for detecting dsRNA. It is believed that one of the primary functions of ADAR1 is to edit endogenous double stranded RNA structures to prevent them being recognized as foreign by the innate immune system.

However, if this hypothesis is correct, it is unclear that ADAR1 is fulfilling this role given that most editing sites have editing levels of less than 1% (Bazak et al. 2014).

Alu-repeats

The targets of ADAR editing are almost all within Alu repeats. These paired elements are composed of two nearby Alu elements, which hybridise to form RNA hairpins, a suitable target for ADAR enzymes. Alu elements are 300 nucleotide long Short-Interspersed Nuclear Repeats (SINEs), a type of retrotransposon. Few Alu elements retain the capacity for transposition.

Retrotransposons are not unique to humans or primates, and yet they have a uniquely high level of RNA editing when compared to the rest of the animal kingdom (with some notable exceptions) (Gommans, Mullen, and Maas 2009). A specific set of circumstances found in humans and primates leads to Alu repeats forming a suitable double-stranded RNA substrate for ADAR activity (Schaffer and Levanon 2021): (1) they are highly abundant, comprising approximately 10% of the human genome, (2) they are concentrated within the introns and 3'-UTRs of mRNAs and (3) they have low divergence compared to LINE elements, which are frequently fragmented. Together, this means there is a high probability of finding two or more oppositely oriented Alu elements within the same pre-mRNA.

8.2.2. Detecting editing sites

Before the advent of Next Generation Sequencing (NGS), the first editing by ADAR enzymes was detected in Xenopus oocytes by thin-layer chromatography (TLC). In their seminal 1988 paper, Bass and Weintraub used TLC to separate and identify radiolabeled mononucleotides derived from RNA after it had been digested by nuclease P1 (B. L. Bass and Weintraub 1988). By comparing the migration of these nucleotides with standards, the presence of inosine was confirmed, demonstrating that the RNA had been chemically modified by the unwinding/modifying activity. Later research used Sanger sequencing to elucidate the

nearest neighbor preferences of ADAR enzymes (Brenda L. Bass 2024; Polson and Bass 1994; Wong, Sato, and Lazinski 2001).

Next generation sequencing (NGS) brought the ability to detect thousands of editing sites across the transcriptome in a single experiment (Shendure and Ji 2008). NGS was widely used from the year 2009 and changed the fields of genomics and transcriptomics. The NGS technologies used to detect RNA edits will be discussed in more detail in Section 8.3. However, despite the increase in sequencing capacity, in humans few additional conserved sites were discovered by NGS (Pinto, Cohen, and Levanon 2014) (Figure 3). The reason for this is that these sites are already highly edited, and it is therefore easier to detect them against background editing. For example, the editing of the GluR2 subunit mRNA which takes place in glutamatergic neurons in close to 100% of the transcripts (Jonas and Burnashev 1995; Hume, Dingledine, and Heinemann 1991). Other examples are the cluster of editing sites in the serotonin receptor 5-HT2CR (Niswender et al. 1999), and the editing event in the voltage-gated potassium channel, KCNA1 (Bhalla et al. 2004).

A-to-I RNA editing levels, are sensitive to RNA sequencing depth, with deeper sequencing revealing many more editing sites, almost all with low editing levels that are often missed at lower coverage (Bazak et al. 2014). Bazak et al. studied two large RNA-seq datasets: the Illumina Human Body Map “HBM” data covering 16 tissues and a deeply sequenced dataset from a Han Chinese individual “YH”. They found that the number of detectable editing sites increases with sequencing depth, with no saturation observed even at a median depth of 5000 reads. Furthermore, they found that as sequencing depth increases, the fraction of Alu repeats showing editing approaches 100%.

Based on these finding they hypothesized that all 761,244 Alu repeats in the human transcriptome had some level of RNA editing. To investigate this, the authors selected 80 Alu elements for ultradeep sequencing. These included 28 Alu elements previously identified as edited in the HBM data and 52 Alu elements that had not shown evidence of editing. The

unedited Alus served as a control to see if deeper sequencing would reveal previously undetected editing events. These sequences were then amplified by PCR from cDNA derived from human brain tissue, which was chosen because of the known high levels of RNA editing in neural tissues, although the levels of editing in the brain turned out not to be exceptionally high compared to other tissues. The amplified Alu elements were then subjected to ultradeep sequencing using an Illumina MiSeq. The sequencing generated extremely high coverage, with most Alu elements covered by thousands of reads (Bazak et al. 2014).

The ultradeep sequencing revealed that virtually all adenosines in the targeted Alu elements (including both strands of the RNA) were edited to some extent, which was projected to bring the total number of editing sites in the human genome to over 100 million. Most of this editing occurs at sites which are rarely edited, and which are in non-coding regions of the transcriptome. Despite the low levels of editing at these sites, the editing in non-coding regions vastly outnumbers the levels of editing at non-synonymous sites in coding regions, which explain only 0.073% of A-to-I editing in brain tissue. The editing levels followed a log-normal distribution, with most sites exhibiting low editing levels but a few showing high levels of editing.

levels of editing at many sites are highly conserved between individuals of the same species.

- In this section, I will focus on de novo detection and quantification of RNA editing sites, where a reliable reference genome is available. Several RNA editing sites databases are available, most notably DARNED (https://darned.ucc.ie), RADAR (http://rnaedit.com/) and REDIportal (http://srv00.recas.ba.infn.it/atlas/ index.html). These databases summarise types of RNA editing events (usually A-to-I or C-to-U), organism and tissue specificity, genomic locations of editing sites, annotation of editing sites, editing frequencies and patterns, comparative and evolutionary information and methodological details (types of sequencing technology and computational methods).

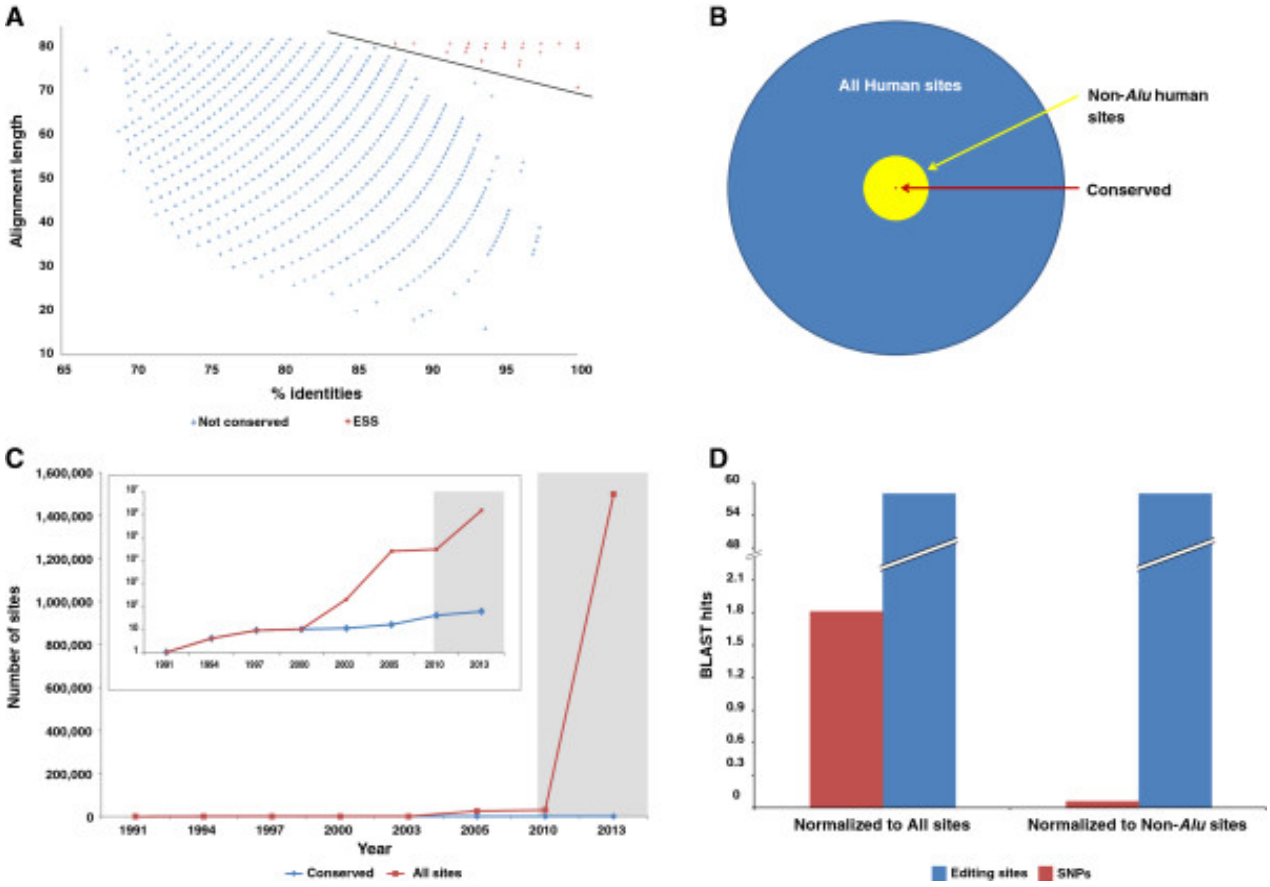


Figure 3 Few evolutionarily conserves sites were discovered by Next generation Sequencing (shared under Creative Commons CC-BY license; (Pinto, Cohen, and Levanon 2014)).

Statistical analysis of RNA editing sites

Identifying RNA editing sites from RNA-seq data (prepared using standard library preparation techniques) involves comparing RNA sequencing data (cDNA) with genomic DNA sequencing data (gDNA) (J. B. Li et al. 2009). The differences observed between the sequences (termed RNA-DNA differences or RDDs) can indicate potential RNA editing events. The challenge associated with differentiating true RDDs from sequencing or mapping errors, or library preparation artifacts necessitates the use of statistical methods which model these potential sources of error.

Legacy methods used for identifying RNA editing sites had significant limitations primarily due to their inability to handle overdispersion, sequencing artifacts, and variability in sequencing data effectively. These shortcomings often resulted in high false-positive rates and reduced the reliability of RNA editing site identification (Piechotta et al. 2017).

- Fixed Threshold-Based Methods: Previously, a common approach involved setting arbitrary thresholds for detecting single nucleotide variants (SNVs). For example, researchers might require a minimum number of variant reads (e.g., 10 reads) or a minimal variant frequency (e.g., 10%) to call an editing site. The use of fixed thresholds is problematic because they do not account for the underlying variability in sequencing depth or quality across different samples or experiments. This leads to both false positives and false negatives.
- Multinomial Distribution-Based Methods: Early approaches to identifying RNA editing sites often relied on statistical models that assumed the nucleotide frequencies at a given site followed a multinomial distribution. The multinomial model is straightforward and computationally efficient, making it an attractive choice for early studies. The primary limitation of the multinomial model is that it often underestimates the variability in nucleotide frequencies observed in sequencing data. This phenomenon, known as overdispersion, leads to an excess of false-positive variant calls. Overdispersion occurs because the actual variance in the data is higher than what the multinomial distribution predicts, causing the model to erroneously identify normal sequencing variability as genuine RNA editing events.

Several software tools, such as SAMtools/BCFtools and REDItools, were designed to detect RNA editing sites. These tools typically employed the above statistical methods, relying on fixed thresholds or simple models. Many early tools did not effectively incorporate replicate sequencing data, which is essential for distinguishing true biological variants from technical

artifacts in noisy or low-depth sequencing data. Additionally, early these tools did not adequately control for sources of error which include (Pinto and Levanon 2019):

- sequencing errors,
- erroneous alignment of sequencing reads to the reference genome,
- biases or errors from library preparation steps,
- somatic mutations,
- differences between the RNA sequencing data and reference genome (aneuploidy is common in laboratory cell lines such as HEK293 (Binz et al. 2019)),
- spontaneous mutations in the RNA or cDNA.

To address these issues, Piechotta et al. developed JACUSA, which employs the Dirichlet-Multinomial distribution, combined with likelihood ratio tests and filtering strategies, to improve the accuracy and robustness of RNA editing site detection (Piechotta et al. 2017).

JACUSA

To address these issues, Piechotta et al. developed JACUSA (and later JACUSA2, which features performance enhancements), a software tool which employs the Dirichlet-Multinomial distribution, combined with likelihood ratio tests and filtering strategies, to improve the accuracy and robustness of RNA editing site detection (Piechotta et al. 2017; 2022). The Dirichlet-Multinomial distribution model accounts for overdispersion and captures the variability in base frequencies across replicates (other currently available tools do not consider data from replicates (Pinto and Levanon 2019)), leading to more accurate identification of true RNA editing events. The parameters of the Dirichlet distribution are estimated using an empirical Bayesian approach, which integrates base call quality scores to adjust the observed data.

In JACUSA, the detection of RNA editing sites involves computing a likelihood ratio based on the Dirichlet-Multinomial distribution. The null hypothesis (H0) posits that the observed data

from two samples originate from the same underlying distribution. The likelihood ratio test evaluates the likelihood that the data from RNA and DNA samples are drawn from different distributions, which would suggest the presence of RNA editing at specific sites. Similarly to RDD detection, RNA-RNA difference (RRD) identification in JACUSA also relies on the Dirichlet-Multinomial model to account for overdispersion. RRD identification is useful for example in comparing experimental conditions.

Piechotta et al. benchmarked JACUSA's performance against other variant callers such as SAMtools/BCFtools, REDItools, and MuTect. The benchmarking process evaluated the true positive rate (TPR), precision, accuracy, and the F1-score (a widely used metric in statistics and machine learning which balances precision and recall). JACUSA has generally shown superior performance, particularly in scenarios involving RNA-RNA comparisons where it consistently outperforms other tools in terms of both precision and recall.

Non-standard library preparation techniques for inosine detection.

Several library preparation techniques have been developed which specifically label inosine residues to enrich these sequences. One such method, iSeq, involves inosine-specific cleavage by RNAse T1 followed by sequencing of the resulting fragments which bear an inosine nucleotide at their terminus (Cattenoz et al. 2013). Another method, ICE-seq (Inosine Chemical Erasing Sequencing), adducts acrylonitrile to the N1 position of inosine to form N1-cyanoethylinosine, resulting in truncation during reverse transcription. By comparing sequences before and after treatment, ICE-seq enables the detection of inosine sites (Sakurai et al. 2014). Additionally, ADAR1 CLIP-seq employs crosslinking immunoprecipitation sequencing to identify the binding regions of the ADAR1 enzyme across the transcriptome, thereby inferring the locations of inosine modifications (Bahn et al. 2015). The newest technique, EndoVIPER-seq uses Endonuclease V (EndoV) in the presence of calcium ions (Ca2+) to bind, but not cleave, inosine-containing RNA fragments. This is critical because in the presence of magnesium ions, EndoV would normally cleave

the RNA. The bound inosine-containing RNA is then pulled down by immunoprecipitation, which employs magnetic beads that are conjugated to an anti-maltose-binding protein (MBP) antibody, targeting an EndoV-MBP fusion protein (Knutson and Heemstra 2020). These methods carry the disadvantage of requiring specialized, experimental and non-commercially available library preparation techniques.

8.3. Next Generation Sequencing

DNA sequencing has undergone significant advancements since its early development in the 1970s, when two-dimensional chromatography was employed for sequence analysis. A breakthrough occurred in 1977 with the introduction of the Sanger chain termination method, which provided researchers with a reliable and reproducible technique for sequencing DNA (Illumina, n.d.). This innovation led to the development of automated sequencing technologies. In 1987, Applied Biosystems released the first sequencing instruments based on capillary electrophoresis. These instruments became the primary tools for the Human Genome Project (1990 – 2004) (Lander et al. 2001).

The release of the Solexa (later Illumina) Genome Analyzer in 2005, after the completion of the Human Genome Project increased sequencing throughput from 84 kb per run to 1 Gb per run (a 10,000-fold improvement). This short-read, massively parallel sequencing method heralded the advent of next-generation sequencing (NGS).

NGS departs from capillary electrophoresis-based sequencing in three ways (Shendure et al. 2017): Instead of conducting reactions in individual tubes, NGS uses a library of DNA templates immobilized on a flat surface, allowing all templates to interact within a single reagent volume. Additionally, in vitro amplification, rather than bacterial cloning, is used to generate copies of each DNA template for sequencing. Finally, rather than determining sequence information by measuring fragment lengths, NGS employs cycles of polymerase-driven incorporation of fluorescently labelled nucleotides—combined with imaging, a process known as "sequencing-by-synthesis" (SBS).

Since its introduction, NGS technology has consistently exceeded the growth predicted by Moore's law, with data output more than doubling each year (“An Introduction to Next-Generation Sequencing Technology,” n.d.).

8.3.1. Illumina

Illumina sequencing is based on SBS. This process involves the stepwise incorporation of fluorescently labelled nucleotides during the synthesis of DNA strands. Genomic DNA or cDNA is fragmented, and adapters are attached to both ends of each fragment (The details of the library preparation vary depending on the quantity and type of nucleic acid starting material and other factors. The details and merits of the library preparations I used in my PhD work are discussed in Section 10.4.3). These adapter-ligated fragments are then amplified through polymerase chain reaction (PCR) and purified to form a sequencing library. The library is loaded onto a flow cell; a pattered glass slide coated with oligonucleotides that are complementary to the adapters on the DNA fragments. Each fragment binds to the flow cell and is amplified into a cluster of identical copies through a process called bridge amplification. During sequencing, each cycle involves the incorporation of a single nucleotide that is labelled with a fluorescent dye whose fluorescence is measured at each cycle (Figure 4). These fluorescent nucleotides are proprietary reversible terminators; after each cycle of imaging, they are enzymatically cleaved, allowing polymerization to continue in the next cycle (Canard and Sarfati 1994; Bentley et al. 2008; Illumina, n.d.).

At each cycle, a laser scans the surface of the flow cell, exciting the fluorophores. The emitted light is detected by a camera and the wavelength (specific to the nucleotide) is mapped to a location on the flow cell. With successive cycles, the sequence of wavelengths mapped to a single location (cluster) on the flow cell can be translated to a DNA sequence (Shendure et al. 2017; Illumina, n.d.).

The process is repeated across millions of clusters simultaneously, enabling the sequencing of large amounts of DNA in parallel, a key feature which enabled the huge increase in throughput delivered by NGS. Paired-end sequencing is an option: both ends of a DNA fragment are sequenced, producing two reads per fragment, one from each end, and doubling the number of reads from a single sequencing run (Illumina, n.d.).

One of the primary advantages of Illumina sequencing is its high accuracy, which is achieved by using reversible terminators. This approach minimizes sequencing errors, particularly in regions of the genome that contain repetitive or homopolymeric sequences, which are often prone to mistakes in other sequencing platforms. Additionally, Illumina sequencing provides exceptionally high throughput, as it can sequence millions of DNA fragments in parallel, producing large quantities of data from a single run (Bentley et al. 2008; Illumina, n.d.).

One notable disadvantage is the short read lengths it produces compared to long-read technologies, such as Pacific Biosciences (PacBio) or Oxford Nanopore. This poses challenges in resolving highly repetitive regions or identifying large structural variants. Additionally, the image capture is inherently delicate. A common problem I encountered is "over-clustering"; too many clusters are generated if too much library is loaded onto the flow cell, making it hard for the camera to discriminate individual clusters, ultimately resulting in a low data output. What's more, the sequencing run cannot be stopped and adjusted once it has started. This is contrast to Nanopore sequencing: devices can be transported in a backpack, library preparation takes a fraction of the time, and relatively 'impure' samples can be sequenced, lending itself to personalized medicine and field research (Wang et al. 2021).

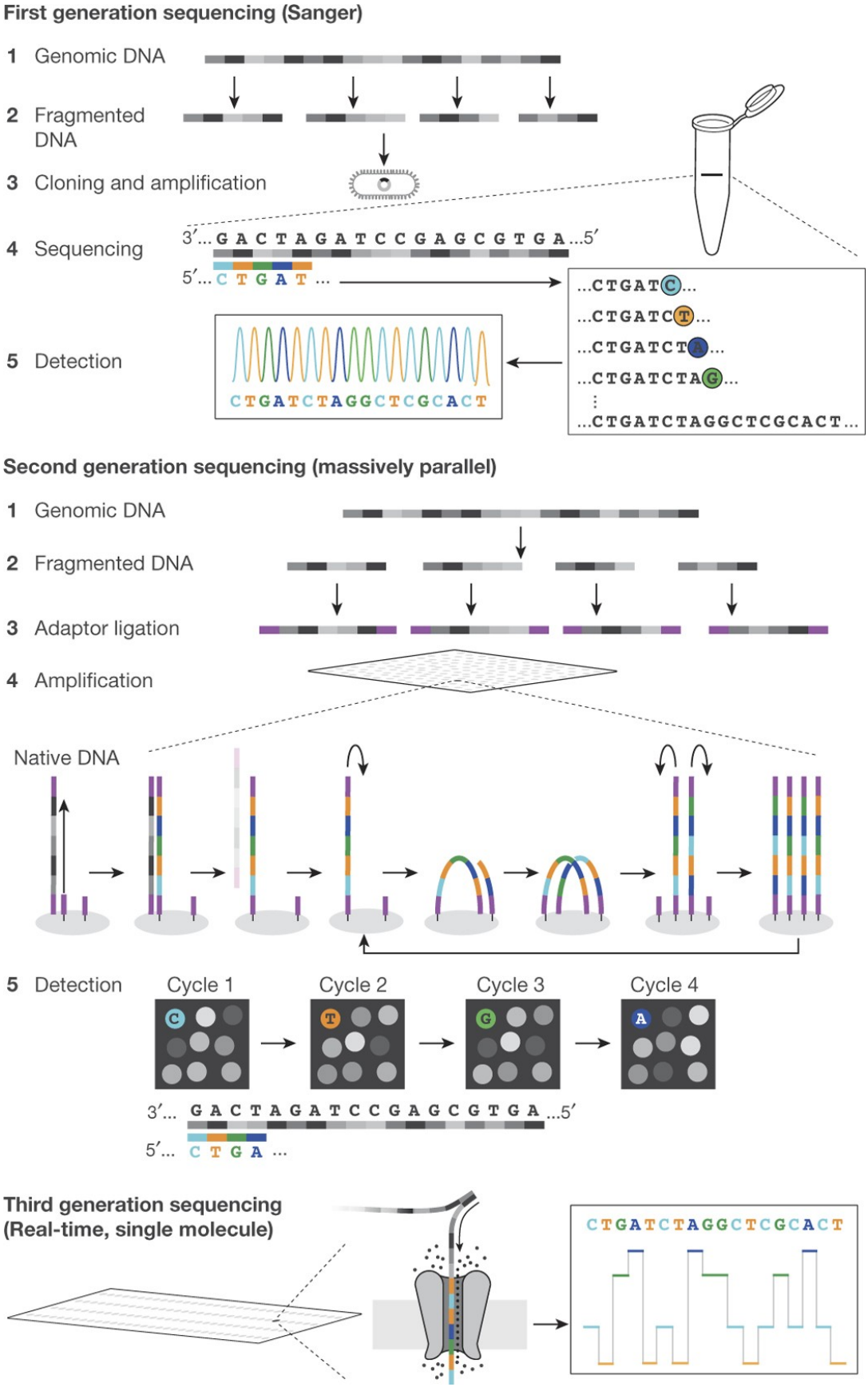

First generation sequencing (Sanger)
1 Genomic DNA
2 Fragmented DNA
3 Cloning and amplification
4 Sequencing
5 Detection
3′... G A C T A G A T C C G A G C G T G A ...5′
5′... C T G A T ...
...CTGATC...
...CTGATCT...
...CTGATCTA...
...CTGATCTAG...
...CTGATCTAGGCTCGCACT...
CTGATCTAGGCTCGCACT
Second generation sequencing (massively parallel)
1 Genomic DNA
2 Fragmented DNA
3 Adaptor ligation
4 Amplification
Native DNA
5 Detection
Cycle 1
Cycle 2
Cycle 3
Cycle 4
3′... G A C T A G A T C C G A G C G T G A ...5′
5′... C T G A ...
Third generation sequencing
(Real-time, single molecule)
CTGATCTAGGCTCGCACT

Figure 4 First, Second (NGS) and Third generations sequencing technologies. (Reprinted with permission; (Shendure et al. 2017)).

8.3.1. PacBio

PacBio was the first technology to achieve highly accurate long reads (>10kb), also known as "HiFi" reads. HiFi reads are comparable in accuracy to Illumina reads (99.8% accurate). This was achieved through the development of Single Molecule Real Time (SMRT) sequencing (Eid et al. 2009) and later, the development of Circular Consensus Sequencing (CCS) (Wenger et al. 2019).

In SMRT sequencing, a single molecule of DNA template-bound Φ29 DNA polymerase is anchored at the base of a Zero-Mode Waveguide (ZMW). The ZMW is illuminated from below by laser light. The ZMW's key purpose is to confine the excitation within an extremely small volume, allowing the detection of individual phospholinked nucleotide substrates against the background of the bulk solution as they are incorporated into the growing DNA strand by the polymerase (Figure 5) (Eid et al. 2009).

CCS generates a consensus sequence from subreads created by several passes of the polymerase around a circular template, thus compensating for the noise and errors present in individual passes. By aligning these multiple subreads, CCS derives a final read, (a HiFi read), that is extremely accurate. The number of passes required to achieve the desired accuracy depends on the length of the DNA fragment (Wenger et al. 2019).

Although much more accurate, the output from a typical PacBio sequencing run is quite modest when compared to the larger Oxford Nanopore sequencers such as the PromethION (see section 10.4.4 for a discussion of the relative merits of PacBio and Oxford Nanopore sequencing).

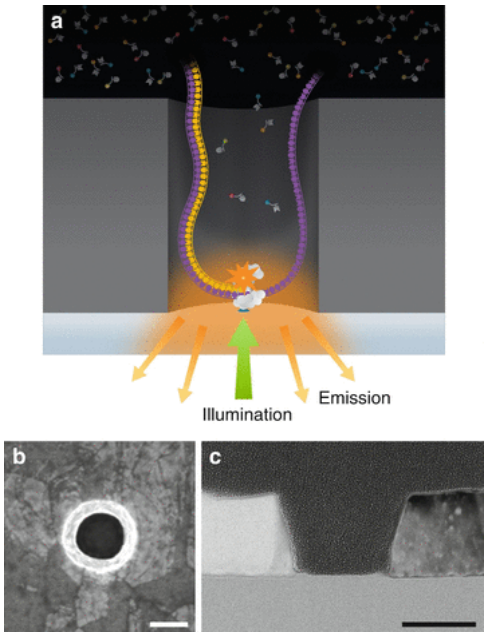


Figure 5 Zero-mode Waveguide (ZMW) ((Eid et al. 2009))

A single Φ29 DNA polymerase synthesising double stranded DNA from a template which anchored at the base of a ZMW. A phospholinked nucleotide associates with the template at the polymerase's active site, leading to an increase in fluorescence in the relevant colour channel. Formation of the phosphodiester bond releases the dye-linker-pyrophosphate product, which then diffuses out of the ZMW, terminating the fluorescence signal. The polymerase then shifts to the next position, where the subsequent nucleotide binds the active site, initiating the next fluorescence pulse.

8.3.2. Oxford Nanopore

Nanopore sequencing works by passing a single-stranded DNA or RNA molecule through a nanoscale pore embedded in an electrically resistant polymer membrane, under the guidance of an electric field. The first commercially available nanopore sequencer was the minION, which was released in 2014 (Jain et al. 2016). Key technical challenges were the engineering of the nanopore and associated motor protein (Dijk et al., n.d.; Maitra, Kim, and Dunbar 2012). The nanopore ('nanoscale protein pore') is an engineered α-hemolysin, from Staphylococcus aureus. A critical improvement to the wild type α-hemolysin was made by Hagan Bayley and colleagues, who engineered a critical transmembrane β-barrel to create a 'recognition site' which enabled characterisation of the mobile DNA sequence in groups of 4 bases (Stoddart et al. 2009). Under the applied voltages required for efficiently capturing the DNA in the nanopores, free moving DNA translocates too quickly for bases to be

characterised. Another critical improvement was the development of a 'motor protein' derived from a DNA polymerase, which allowed ssDNA to be "ratcheted" through the pore one base at a time (Cockroft et al. 2008).

As the DNA or RNA translocates through the pore, the unique structure of each 4-nucleotide quad causes a characteristic disruption in the ionic current. These current changes are detected and recorded. Each channel is associated with a single electrode and application-specific integration circuit (ASIC), which controls the voltage across the pore. The ASIC allows sequencing to be started and stopped and the voltage can be reversed if pores become blocked. The recorded current fluctuations are processed 'on the fly' by recurrent neural networks that decode the sequence of nucleotides. This allows live feedback of sequence quality, the presence of sequences of interest and the enrichment of sequences of interest.

Unlike traditional sequencing methods that require amplification or synthesis, nanopore sequencing allows real-time sequencing of long DNA or RNA molecules, providing the advantage of direct readout of native molecules, including modified bases. Although much progress has been made with direct RNA sequencing using nanopore, it was low accuracy when I began my PhD and still trails DNA sequencing on the Nanopore platform (Jain et al. 2022). One disadvantage of SBS is that errors which are introduced by the polymerase during sequencing are not quantified by the manufacturer reported quality scores. Because Nanopore (unlike Illumina and PacBio) uses a direct sequencing approach, these errors are avoided (Sun et al. 2024). See section 10.4.4 for a discussion of how quality scores are calculated. This project benefitted from continuous improvements in engineering of the α-hemolysin nanopore, motor protein and basecalling algorithms that came with the R10.4.1 flow cells and V14 chemistry (Ni et al. 2023; Bogaerts et al. 2024).

Nanopore sequencing requires unique bioinformatics approaches due to the nature of its long reads, high error rates, and real-time data generation. These requirements include

specialized tools for base-calling, error correction, and read alignment (Wang et al. 2021; Li 2018).

9. Terminology

In the results chapters of this work, I use various Timestamps-specific nomenclature, which I explain in the table below.

Table 1 Terminology used in this work.

Calibration experiment	Cells are lysed at regular intervals following treatment with actinomycin D. This allows one to observe the change in editing occurring at editing sites after a known increment in time. Calibration experiments serve two purposes: (i) To identify fit sites and (ii) The editing rate parameters in Equation 2 are fit to the editing changes observed at fit sites in calibration experiments. These parameters are then used to infer time in discovery experiments. In this work, the HEK293 and human iPSC cortex calibrations are combined to achieve a larger number of editing sites with known editing rate parameters (See 10.4.7.1)
Discovery experiment	In these experiments, cells were perturbed and then sampled at fixed points thereafter. The calibrations editing rates are then used to calculate the ages of mRNA molecules. (e.g. the human monocyte LPS experiment - Section 12)
Editing sites	Sites in the transcriptome at which A-to-I editing is detected by JACUSA.
Fit site	The subset of editing sites which are well fit by an exponential CDF
Per-site age	The estimated age from a single fit site.
Gene age	The mean age of the transcripts produced from a single genomic locus (e.g. all the GAPDH mRNAs). Calculated by taking the mean of the per-site age estimates over all the sites annotated to that gene.
Single molecule age	The estimated age of a single transcript (single molecule of mRNA). Calculate by the Maximum Likelihood Estimate (MLE)

Maximum Likelihood Estimate (MLE)	From long-read sequencing .bam files, base calls at fit sites are extracted. The age of individual transcripts is modelled by constructing a likelihood function for the age of the transcript (see Section 10.4.7.3).

10. Endogenous editing of mRNA in human cells

The transcriptome is a dynamic population of molecules whose composition reflects a cell's internal state. RNA sequencing provides detailed information about the instantaneous composition of a cell's transcriptome, but methods to recover temporal information by sequencing have so far been limited (Choi et al. 2022).

Previously, Rodriques et al. showed that the ages of individual RNA molecules can be recorded with hour-scale fidelity by the activity of an engineered adenosine deaminase on an engineered RNA substrate. In principle, this method can be used to infer transcriptional dynamics non-invasively with hour-scale time resolution from sequencing data collected at an endpoint. Like other molecular recorders, this method requires significant genetic engineering, as it depends on both an engineered deaminase and an engineered RNA substrate. However, as I discussed in the previous section, A-to-I editing catalysed by the ADAR enzymes is virtually ubiquitous in human tissues (Tan et al. 2017) and is widespread in the human transcriptome (Bazak et al. 2014), possibly providing us with an endogenous cellular RNA timestamp.

Here, I hypothesise that the Timestamps concept could be extended to endogenous, un-engineered transcripts in un-engineered human cells by performing deep and full-length RNA sequencing and using endogenous ADAR activity to approximate the age of single molecules of mRNA.

10.1. Interval sequencing (calibration) of human iPSC-derived neurons

Adenosine ribonucleotides are deaminated by ADAR enzymes in the nucleus and cytoplasm. Rodriques et al. found that RNA editing sites were suitable for inferring the age of RNAs if the editing activity at those sites was well-fit by an exponential cumulative distribution function (CDF). Several criteria must be met for endogenous editing to provide a

suitable timestamp. RNA edits should accumulate with time from transcription in a relatively predictable manner, so that the dynamics of edit accumulation can be described mathematically. Ideally, editing at one site would not have a large effect on editing at other sites The accumulation of edits should happen on a timescale which is useful for measuring transcriptional dynamics on the timescales at which cellular processes happen (i.e. minutes and hours). Additionally, these conditions should be met by a sufficiently broad portion of the transcriptome to have general purpose use (i.e. not measuring only a small subset of genes).

To measure whether endogenous RNA editing is suitable for inferring RNA age, I designed an experiment which would allow us to observe the accumulation of RNA edits across the transcriptome at regular intervals. Human iPSC-derived cortical and midbrain neurons were treated with actinomycin D (ActD; an inhibitor of RNA polymerase II), thus preventing the creation of new mRNA. At fixed timepoints thereafter, I lysed the wells in duplicate and extracted the mRNA for deep, full-length, short-read RNA sequencing (Illumina), thereby obtaining mRNA with a known increment in age (Figure 6).

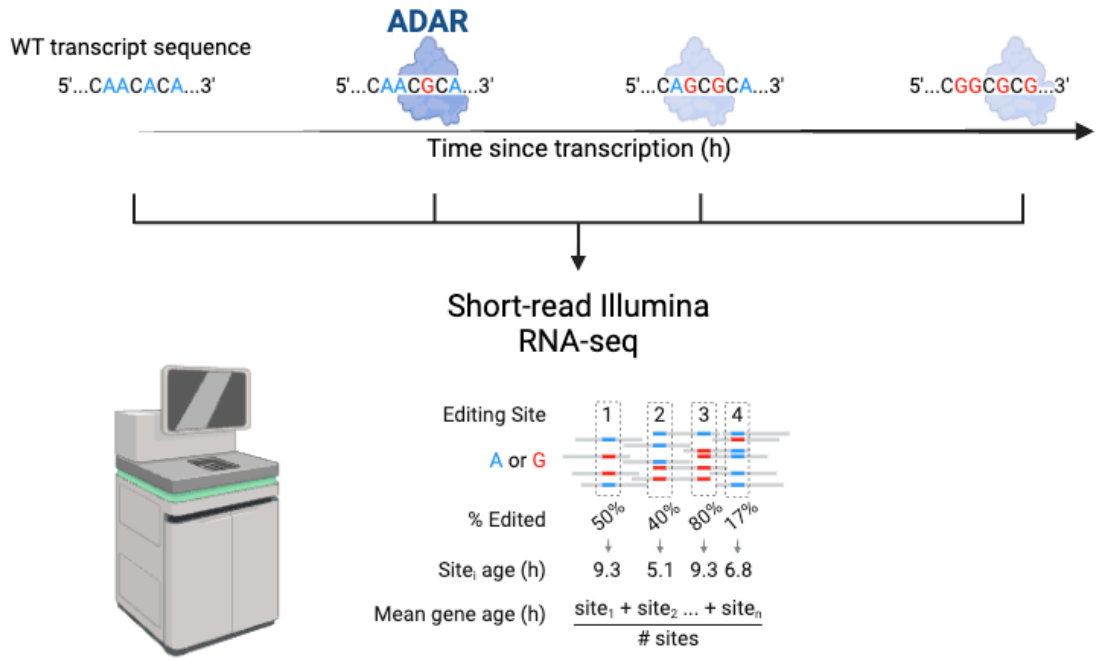


Figure 6 Interval sequencing (calibration) of iPSC-derived neurons treated with actinomycin D.

The library preparation protocol used ensured that I obtained stranded (strand specific) RNA sequencing data (Methods for Chapter 10). For these experiments, it was not important to discriminate between A-to-I editing resulting from the two isoforms of ADAR.

After trimming primer sequences, aligning to the human reference genome and identifying RNA edits using JACUSA2 (10.4.6.1), I was able to study the editing events that had taken place at each interval. Reassuringly, I found that A-to-I edits are the most common in these human cells lines, an indication that I are identifying real ADAR activity (Figure 7A; cf. Figure 1 of Pinto and Levanon 2019). To my surprise, I found that RNA edits accumulated across the transcriptome after actinomycin D addition (Figure 7B). This was the first indication that my approach might work. Editing events were spread across the transcriptome in an approximately log normal distribution, which has been reported by others (Bazak et al. 2014). Most genes contain a small number of putative editing sites, with a long tail of genes with many putative editing sites (Figure 7C).

I found that 1,044,970 adenosines are significantly differentially edited in hiPSC-derived cortical neurons 8 hours after adding ActD. Of those editing sites, 162,741 (15.6%) are well fit by an exponential CDF (Equation 1; Table 4 - Appendix). The editing fractions at five editing sites in the 3'UTR of the *EIF3M* gene are shown at 2-hour intervals following ActD treatment. The fit of Equation 1 is shown by the black line (Figure 8A). Using the fitted per-site editing rates and the observed editing fractions allowed us to estimate the ages of genes across the transcriptome (here, mean gene ages are calculated by averaging the per-site ages; Equation 2). "Genes" encompasses all the mRNAs which are transcribed from a given genomic locus i.e. an ensemble of mRNA molecules of the same species e.g. all GAPDH mRNAs. The average age of the observable transcriptome increases linearly from addition of ActD. The differences in mean gene age are approximately equal to the intervals between timepoints in the calibration experiment (i.e. the gradient of the increase is 1). Note that genes begin the calibration experiment with a steady state level of editing.

Unsurprisingly, I found that editing sites which are infrequently edited did not make good timestamps. Additionally, I found that editing sites in introns (which unfortunately comprise most editing sites in the neuron transcriptome) were also poorly fit by a CDF. The likely explanation for this is that these editing sites are frequently spliced out of the pre-mRNA, whereas the sites in 3'UTR are less frequently spliced.

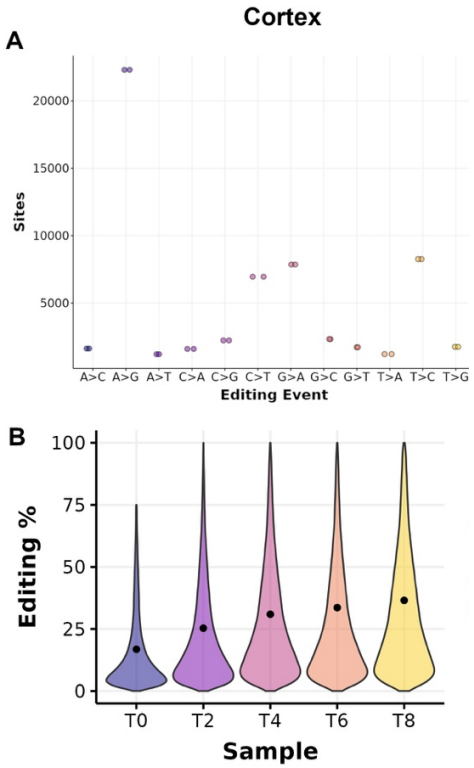


Figure 7 RNA editing in human iPSC-derived neurons

(A) Numbers of all editing types detected by JACUSA.
(B) Violin plots showing the faction of detected sites which are edited at each time point.

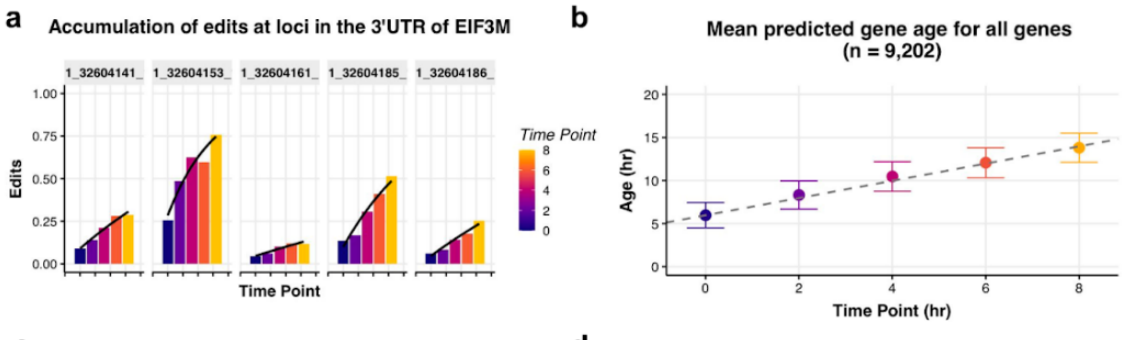


Figure 8 Accumulation of edits in human iPSC-derived cortical neurons (analysis performed by James Bayne)

Prediction of transcript ages from endogenous ADAR-mediated A-to-I editing in human cortical neuron culture (n=2). (A) Time-dependent accumulation of edits at selected sites on EIF3M from short-read sequencing with editing model fit overlayed (black curve). (B) Mean gene ages for all genes

for which mean ages could be calculated. Error bars show the 95% CI. The dashed line has a gradient of 1.

10.1. Determination of single-molecule age

So far, I have determined gene ages; the mean ages of an ensemble of mRNAs derived from the same gene. However, determining the ages of single molecules of RNA would allow population age-distributions to be constructed for each gene. Of course, this would be restricted to genes which had sufficient editing in the calibration experiment (Section 10.4.7.1). I believed these age distributions would encode past transcriptional activity. If all a transcript's editing sites could be captured in a single long read, I reasoned that Timestamps could provide an estimate of the age of single molecules of mRNA.

When I started the long read sequencing for this project, the ONT R10 flow cells were not yet available and PacBio circular consensus sequencing was far ahead in terms of mean per-base accuracy (J. Sun et al. 2023; Bogaerts et al. 2024; Ni et al. 2023) (See Section 10.4.4 for a comparison of these technologies). Therefore, proceeding with PacBio, I performed long-read sequencing on the 0-hour and 8-hour timepoints from the cortex experiment.

When compared to the Illumina reads, I found that each read in the PacBio data captures more editing sites on average (Figure 9). This suggested that for several hundred genes, using long-read sequencing was indeed capturing many more edits per read than Illumina sequencing. The genes are likely to be the ones with multiple Aliu-repeats and therefore many editing sites spread hundreds of base pairs apart.

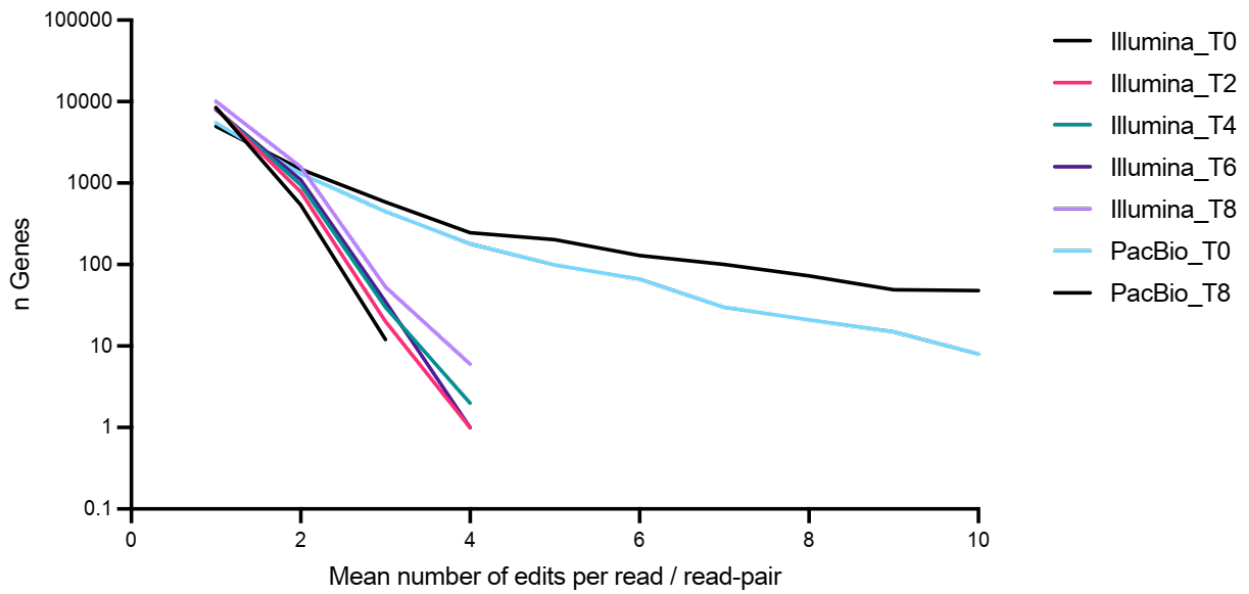
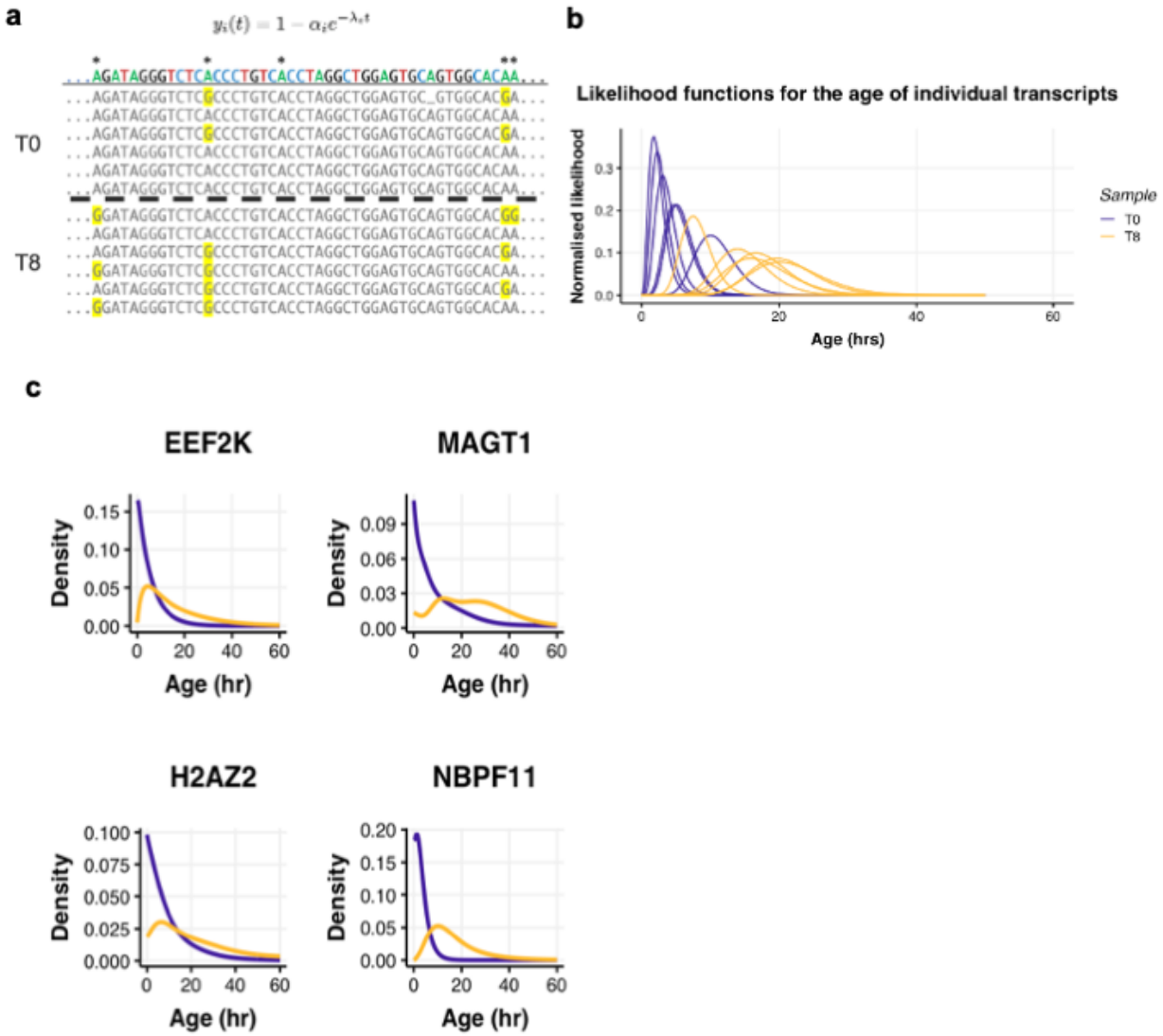


Figure 9 Number of edits per read captures by Illumina and PacBio sequencing (analysis performed by George Young (Bioinformatics – Francis Crick Institute).

Individual reads from the PacBio sequencing which mapped to the 3' UTR of the EIF3M gene are shown in Figure 10C. As expected, the reads are more edited 8 hours after ActD addition. The corresponding transcript likelihoods show which transcript ages are most likely to explain the observed editing (Figure 10D). The peak of the likelihood curves represents the most likely age of this individual molecule of mRNA i.e. the per-transcript Maximum Likelihood Estimate (MLE); Equation 3). The likelihood functions of individual mRNA transcripts are summed and normalised to give per-gene likelihood functions. The per-gene likelihood functions of four different genes are shown in Figure 10E.

The distribution of transcript MLEs is shown in the violin plots in Figure 10F. Encouragingly, the incremental change in MLE age for various genes corresponds to the increment between timepoints in the calibration experiment (Figure 10F – labels). I found that transcript MLEs agree with the population mean ages from the per-site data, suggesting that Timestamps accurately determines the age of single mRNA molecules (Figure 10F – white circles).


a
$y_i(t) = 1 - \alpha_i e^{-\lambda_i t}$
T0
T8
b
Likelihood functions for the age of individual transcripts
Normalised likelihood
Age (hrs)
Sample
T0
T8
c
EEF2K
MAGT1
H2AZ2
NBPF11
Density
Age (hr)

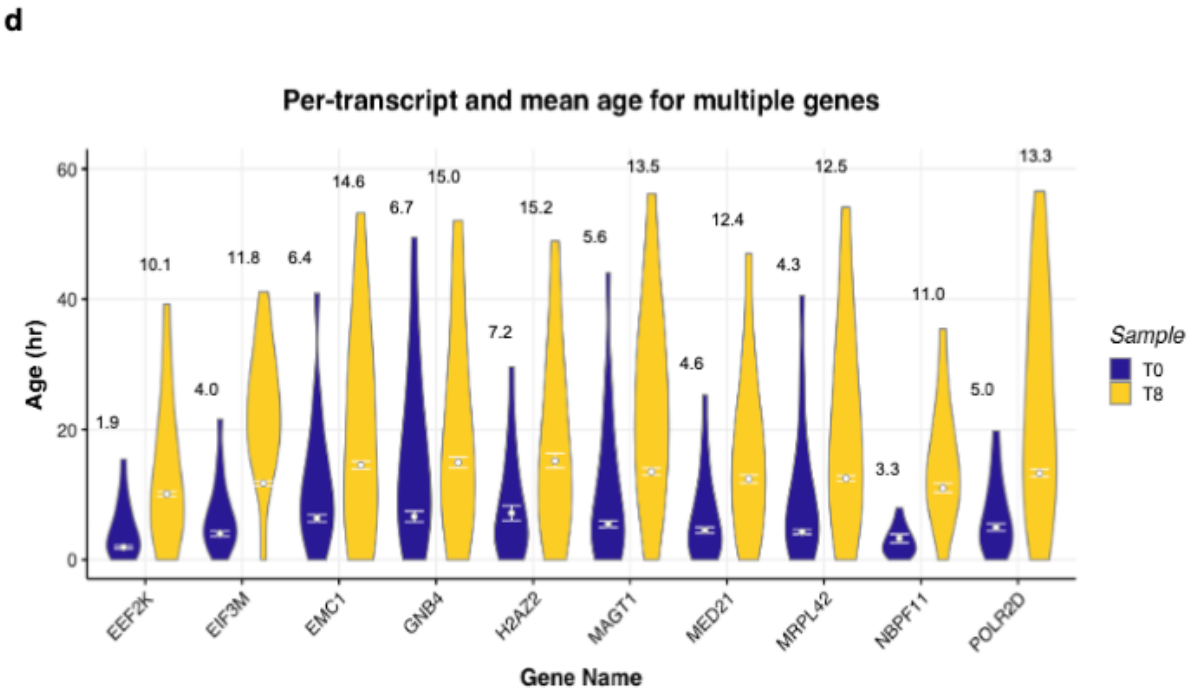


Figure 10 Long-read sequencing of human IPSC-derived cortical neurons (analysis performed by James Bayne)

(A) Individual reads of EIF3M mRNA from long-read sequencing of the calibration dataset. Asterisks denote the sites shown in a and highlights denote editing events.

(B) The likelihood functions of the age of the individual reads shown in panel C, normalised by area.

(C) Distribution of the age of the transcript pool of select genes, calculated by summing the per-transcript likelihoods.

(D) Combination of both per-transcript and per-gene ages. The distribution of age of the transcript pool from long-read sequencing are shown as violins. The mean age (determined from short read sequencing) is shown as points with the standard error shown as error bars (mean age in hours is shown above each violin to 1 decimal place).

10.2. Development of a promiscuous, hyperactive RNA editor – Generation 1

Although the results from the human iPSC-derived neurons were encouraging, I sought to increase the editing levels across the transcriptome so that genes with low levels of endogenous ADAR activity might be studied (for example, mouse cells which have much

lower levels of editing than human cells due their lack of non-divergent repeats i.e. Alu repeats (Neeman et al. 2006)). While those developing base editors for therapeutic purposes have sought to develop mutant base editors which are hyperactive and specific for a target sequence, I sought to develop base editors which are hyperactive and have loose sequence requirements, thereby increasing the number and breadth of A-to-I edits.

10.2.1. Identifying a suitable catalytic domain mutant

I cloned six constructs whose activity has previously been characterized. These were the full-length protein and catalytic-domain (CD) only versions of three ADAR2 mutants: wild type, E488Q (putative hyperactive), and T490A-E488Q (putative inactive mutant) mutants. These constructs were expressed under a strong promoter (CAG) and transfected into HEK293T cells. Twenty-four hours after transfection, ActD was added to the cells to pause transcription and then the cells were lysed at 0 and 8 hours. Figure 12 shows that the full-length ADAR2(E488Q) mutant "AG032" produced many more editing sites than the other candidates at the 8-hour timepoint. Interestingly, the CD-only construct produced far fewer edits than the full-length version, pointing to the importance of the dsRNA binding domains in

Note: by sequencing only one timepoint, I could only study instantaneous levels of editing at 8 hours. I could not therefore calculate editing rate enhancements by the base editors. This shortcut, which was designed to allow us to sequence this experiment on a single Illumina High-Output MiniSeq run (50 million paired-end reads) would turn out to be a error (Figure 12D).

10.2.2. Retargeting the catalytic domain mutant

Sequence preferences for ADAR1 and ADAR2 as mostly determined by the CD's substrate binding preferences (Wang, Park, and Beal 2018; Polson and Bass 1994; Kuttan and Bass

2012). I reasoned that (i) using a hyperactive mutant of the CD with relaxed sequence preference (i.e. ADAR2(E488Q)) and (ii) retargeting this mutant CD to non-Alu targets by swapping out the dsRNA binding domains for alternative targeting domains would enhance non-Alu ADAR activity.

One strategy is to retarget the activity of mutant ADAR CD using an antisense guide RNA. I explored the possibility of targeting ADAR activity to the poly-A tail, thereby concentrating mutant ADAR in the vicinity of the mRNAs. Poly-A tails are a feature of almost all eukaryotic mRNAs, except for certain histone transcripts. These tails, added during transcription, are required for the export of mature mRNAs to the cytoplasm. In mammals, poly-A tails typically measure around 200 nucleotides (Passmore and Coller 2022). While edits deep within the poly-A tail would be difficult to map to the reference genome without long-read sequencing, I believed that many edits would be within the 3'UTR due to its proximity to the poly-A tail.

Montiel-Gonzalez et al. has created a construct whereby the ADAR2(E488Q) CD is fused to the phage lambda N-protein (Nlambda-ADAR; Figure 11) (Montiel-Gonźalez, Vallecillo-Viejo, and Rosenthal 2016). In the lambda phage, N protein has a high affinity for the BOXB RNA motif which it uses to localise to *nut* (N utilization) sites that are in the early operons of the phage genome. Once bound, the N protein interacts with the host's RNA polymerase and other host factors to modify the polymerase, allowing it to ignore downstream termination signals (Lazinski, Grzadzielska, and Das 1989).

In the Nlambda-ADAR construct, the N-protein effectively tethers ADAR to a BOXB-containing guide RNA. By co-transfecting the Nlambda-ADAR construct with a poly-U guide RNA containing the BOXB motif, I aimed to target ADAR to the poly-A tail of mRNAs. One challenge with this approach would be to express the poly-U guide RNA from a plasmid. This is because poly-T DNA sequences efficiently terminate eukaryotic RNA polymerase III (Gao, Herrera-Carrillo, and Berkhout 2018). For initial *in vitro* experiments, it would suffice to transfect cells with the guide RNA directly, without a DNA intermediate.

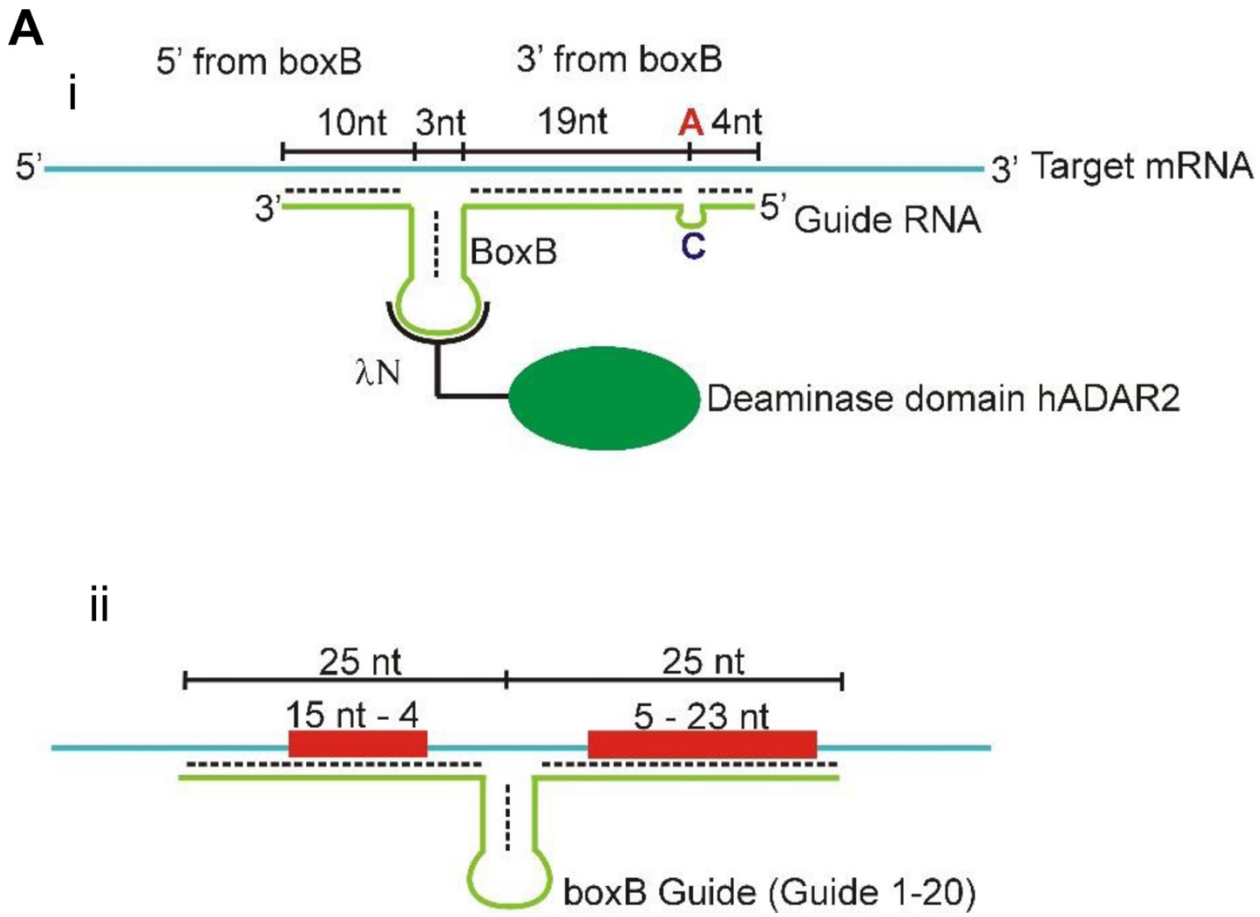


Figure 11 the Nlamdba-ADAR(E488Q) construct (Reprinted with permission from (Montiel-Gonźalez, Vallecillo-Viejo, and Rosenthal 2016))

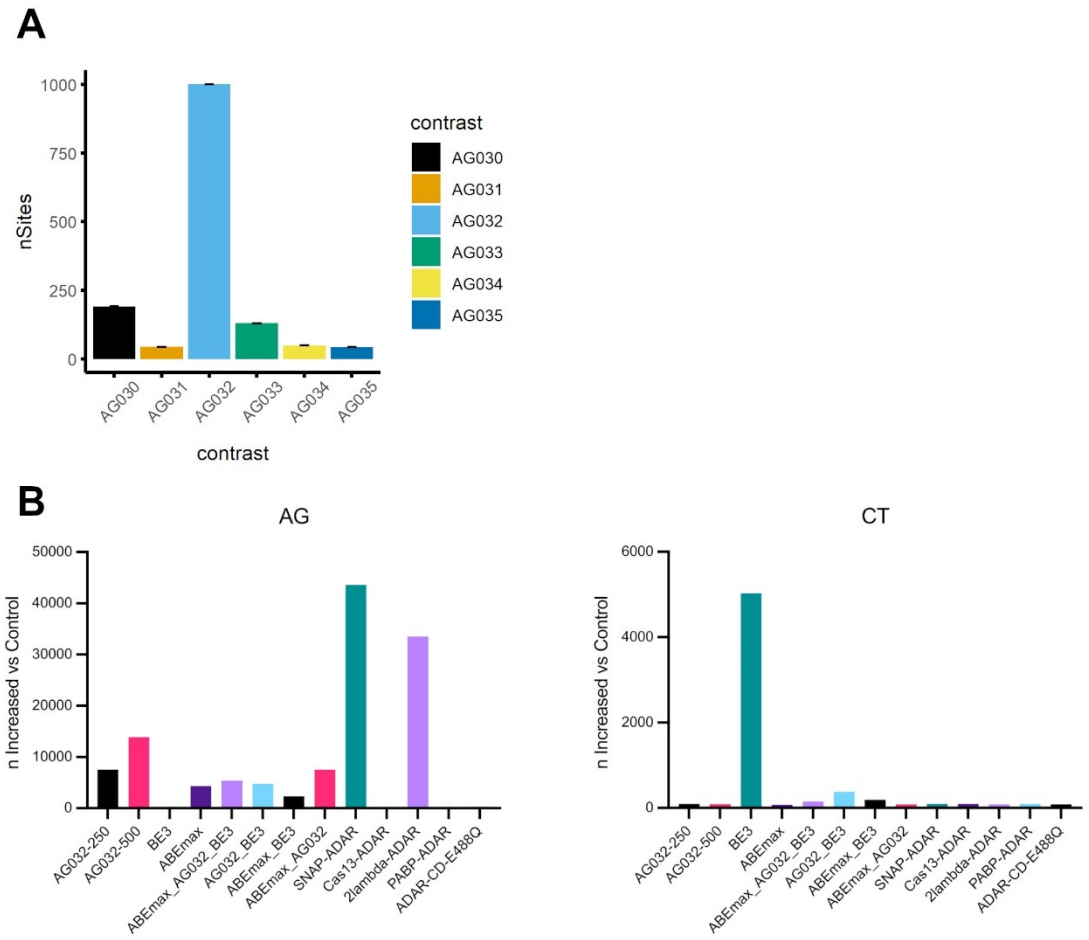


Figure 12: Searching for a promiscuous, hyperactive RNA base editor

(A) number of sites which were differentially edited for the six ADAR2 mutants as compared to empty vector control detected by JACUSA with Z>1.96. (AG030/031 = full length and catalytic domain T490 ADAR2 mutant, AG032/033 = full length and catalytic domain E488Q ADAR2 mutants, AG034/035 = full length and catalytic domain E488Q-T490A ADAR2 mutants.

(B) number of sites which were differentially edited for poly-A targeted ADAR2 constructs as compared to empty vector control. Shown are counts of A-to-G (left) and C-to-T transitions (right) detected by JACUSA2 in differential mode relative to the control (empty plasmid). AG032 = ADAR2(E488Q); AG032-250 = 250ug per well; ADAR-CD-E488Q = ADAR2(E488Q) catalytic domain only; ABEmax_AG032_BE3 is an equimolar mixture of three plasmids. Other plasmids are described in 10.4.5.

Other editors in this poly-A targeting class of constructs included: (i) ADAR2 fused to poly-A binding protein, (ii) Cas13 (co-transfected with poly-A targeting gRNA) (Cox et al. 2017) and

(iii) a SNAP-tag mutant that binds with high affinity to gRNAs containing benzylguanine-modified residues (Vogel et al. 2018).

10.2.3. Cytosine Base Editors (CBEs) and ABEmax

Inspired by the findings of the Joung lab; that cytosine base editors (CBEs) fused to a Cas9 with nickase activity induce off-target edits throughout the transcriptome (Grünewald, Zhou, Garcia, et al. 2019), I sought to find out if these editing constructs, originally intended for therapeutic purposes, could be repurposed for timestamping the human transcriptome. Specifically, I wanted to compare their editing activities to that of the E488Q ADAR2 mutant and to look at whether combinations of the CBEs and ADAR would induce high levels of edits.

In gene therapy, the cytosine deaminase APOBEC1 has been repurposed for editing the genome, an application in which it functions as a DNA cytosine deaminase. This means that APOBEC1 can deaminate cytosine (C) bases in DNA, converting them to uracil (U), which can then be replicated as thymine (T) in DNA sequences, leading to C-to-T edits.

In nature, however, the primary function of APOBEC1, is to deaminate cytosines in RNA, not DNA (Salter, Bennett, and Smith 2016). This RNA editing capability was first observed in the context of apolipoprotein B (APOB) mRNA, where APOBEC1specifically targets cytosine 6666 in the rat apoB mRNA, converting it into uracil. This conversion changes the codon at position 2153 from a glutamine codon (CAA) to a stop codon (UAA). The introduction of this premature stop codon results in the production of a truncated protein, apoB48, which is about 48% of the length of the full-length apoB100 protein (Boström et al. 1990; Lau et al. 1990). Subsequent research showed that APOBEC1 could modify cytosines in multiple RNA transcripts beyond just APOB (Rosenberg et al. 2011). This broader RNA editing activity raised concerns about the specificity and potential off-target effects when using APOBEC1 in cytosine base editors. The ability of APOBEC1 to edit cytosines in RNA meant that its overexpression for therapeutic purposes could have off-target activity in the transcriptome – activity which was detected by Grünewald et al. (Grünewald, Zhou, Garcia, et al. 2019).

ABEmax, an E.coli TadA optimised for editing of single stranded genomic DNA (Gaudelli et al. 2017) was also noted by Grünewald et al. to have widespread off-target activity, and so I also studies this construct.

10.2.4. Construct comparison experiment

I transfected plasmids and (guide RNAs, if required) for all of the above constructs into HEK293 cells plated in well plates in triplicate. Twenty-four hours later, I added ActD to pause transcription and then lysed the cells at 0 and 8 hours before preparing only the 8-hour mRNA for Illumina sequencing on a NextSeq 550 with a mid-output kit (approximately 260M paired-end reads). *(Note: again, sequencing only the 8-hour timepoint was designed to save space on the sequencer, as I believed that measuring the instantaneous levels of editing would be sufficient)*.

Two clear winners emerged: SNAP-ADAR and Nlambda-ADAR, both poly-A targeted constructs. Both constructs showed several-fold higher levels of instantaneous editing than their competitors. Although SNAP-ADAR showed slightly higher levels of editing than Nlambda-ADAR, I dropped it from further development because its guide RNA cannot be genetically encoded due to the need for an O6-benzylguanine base modification. The ability to express the guide RNA from a plasmid or viral vector is essential for *in vivo* use (e.g. in mice).

10.2.5. Assessing Nlambda-ADARs guide RNA requirement

To my surprise, I found that Nlambda-ADAR edits were not concentrated in 3' UTRs, but found throughout transcripts. This raised my suspicion that the guide RNA was not necessary and perhaps the Nlambda-ADAR construct was binding directly to mRNA through non-BOXB binding of the lambda N-protein. I therefore performed another experiment with three conditions in triplicate: HEK293 cells were transfected with Nlambda-ADAR plus poly-T guide RNA, Nlamdbda-ADAR only, or with an empty vector (negative control). As previously, cells were treated with ActD and only mRNA from the 8-hour timepoint was sequenced on a NextSeq 550.

I found that indeed, the guide RNA made little difference to the number of RNA edits (a mean of 38616 versus 36153 editing sites found by JACUSA in the guide RNA versus no guide RNA conditions, respectively; n=3). Additionally, I was able to perform differential gene expression analysis between the Nlambda-ADAR and empty vector cells, to detect any cellular perturbations as a result of the high Nlamdba-ADAR expression. Reassuringly, I did not find any significantly differentially expressed genes between then Nlambda-ADAR and empty vector controls.

10.2.6. Interval sequencing (calibration) of HEK293 cells transfected with Gen 1 editor

After identifying a potential transcriptome-wide editor, I performed a new calibration experiment to characterise its performance at timestamping the transcriptome.

I transfected HEK293 cells with the best-performing Gen1 editor – Nlambda-ADAR (excluding SNAP-ADAR because its guide RNA could not be genetically encoded). I did not transfect the guide RNA for Nlambda-ADAR because my previous experiments had shown this was not necessary for its transcriptome-wide editing activity.

I carried out the same calibration experiment as in Figure 6, this time with 13 time points in triplicate. As for the interval sequencing of the cortical neurons, I performed deep, full-length RNA sequencing with Illumina on the HEK293 cells to identify editing sites and fit the editing rate model in Equation 1. I then used these editing rate parameters to calculate mean per-site (Figure 14) and then mean gene age (Figure 15). Conceptually, reads from older transcripts will be more highly edited than those form younger transcripts. Finally, I performed long-read sequencing on a sub-set of timepoints from the experiment and verified that the single molecule age predictions matched ground truth. This process is outlined in the schematic in Figure 13A.

In HEK293 cells, I detected 428,148 significantly differentially edited sites log-normally distributed across the transcriptome (Figure 13B). Of these, 68,599 sites (16.0%) across 6386 genes were fit well by an exponential CDF (Figure 16A). Figure 13C shows the fit line of Equation 1 approximates well the accumulation of edits at 4 sites in the GATC gene in HEK293 cells. The mean age of a gene (a gene's transcripts) can be calculated taking the average age of all the fit sites mapping to those transcripts. Figure 13D shows the average age of the measurable transcriptome (n=6386 genes) increases linearly with time from ActD addition at t=0. Here, the delta age is plotted, so that the origin is at (0,0). Furthermore, I found the editing rates between HEK293 cells and iPSC-derived cortical neurons to be well correlated (Pearson Correlation = 0.69; Figure 16B).

Next, I performed long-read sequencing on mRNA from the 0, 4, 8, and 16H timepoints using the PacBio platform (Figure 13E-G). The cortex experiment showed that long read sequencing allows us to infer the ages of individual mRNA transcripts by observing the editing states of all editing sites on a single molecule. Figure 13E shows likelihood functions for individual transcripts of the GATC gene, at each of the observed timepoints. The peak of the likelihood function corresponds to the age which is most likely to explain the observed editing (the MLE). Young transcripts (from the earlier timepoints) produce narrow and sharp peaks indicating greater certainty in the predicted age. Older transcripts produce broader likelihood functions, a result of dispersion of the possible editing combinations on the molecule due to stochastic editing.

The MLEs can be extracted from each transcript likelihood function and plotted as histogram which represents all the transcript ages in the HEK293 cells. Figure 13F shows that as I move from newer to older experimental timepoints (blue to yellow) the peak of the distribution shifts from left to right as expected. Encouragingly, the difference between the mean ages of the transcripts (μ) are approximately equal to the epoch between experimental timepoints (Figure 13F). Furthermore, the mean ages of various genes (calculated from the short read

sequencing data) correspond well with the mean ages calculated from the single molecule MLEs. This shows correspondence between the short read and long read sequencing results. However, the distribution of MLE is much wider than the distribution of mean age estimates, particularly for the older timepoints. Dispersion is seen in both the Illumina and PacBio data but is more pronounced in PacBio due to the much lower sequencing depth achieved with PacBio as compared to Illumina. The low throughout of the PacBio Sequel II system, and the release of Oxford Nanopore R10 flow cells ultimately motivated us to switch to nanopore sequencing (see discussion in 10.4.4).

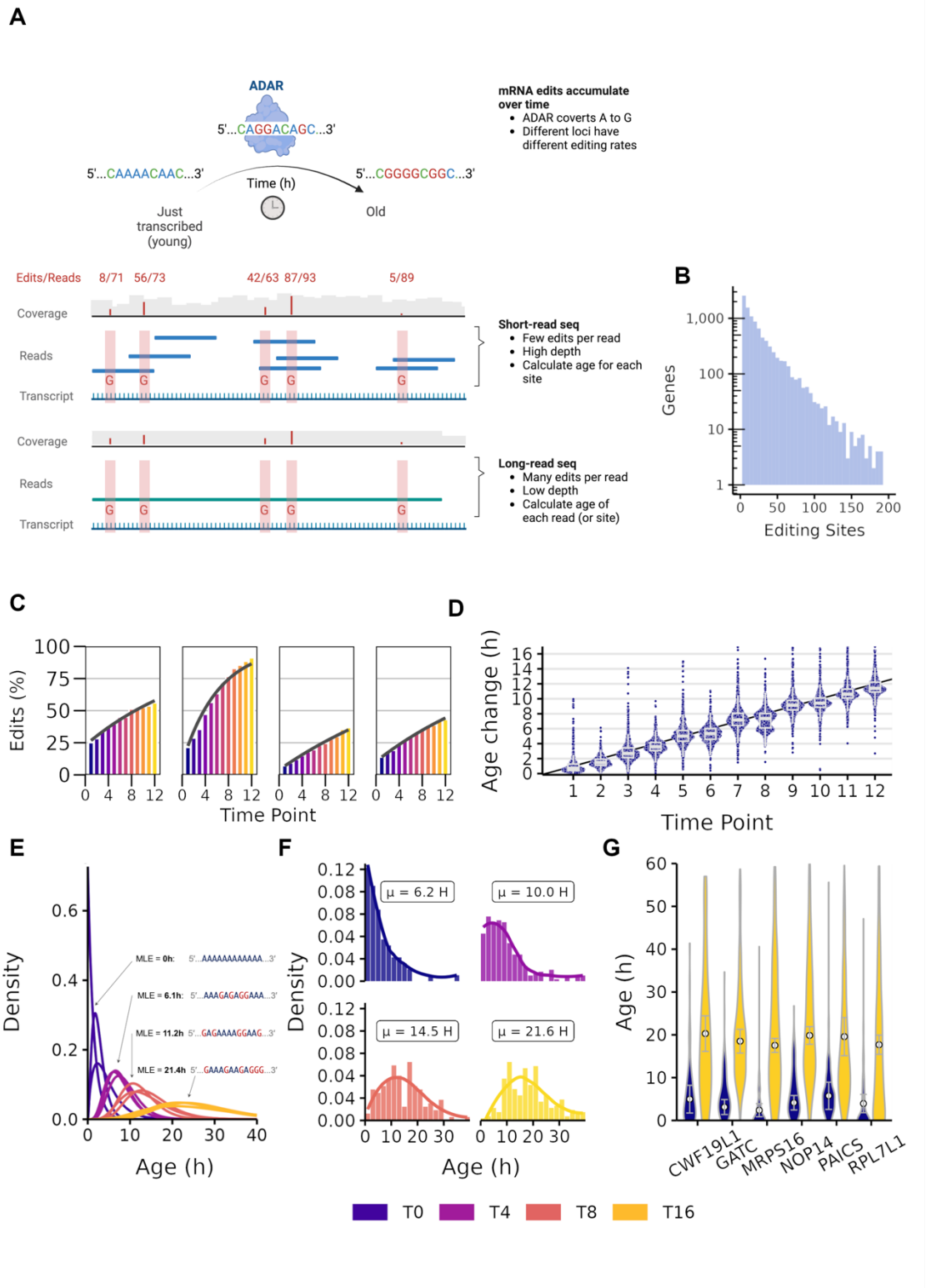


Figure 13 RNA editing encodes the age of human transcripts in HEK293 cells expressing a hyperactive, promiscuous editor

(A) Prediction of transcript ages from endogenous ADAR-mediated A-to-I editing in HEK293 cells transfected with Nlamdba-ADAR (n=3).

(B) Histogram showing distribution of edits per gene across the transcriptome.

(C) Time-dependant accumulation of edits at four sites on GATC from short-read sequencing with editing model fit overlayed (black curve).

(D) The mean age of all genes is calculated from the short-read HEK calibration data and the time change compared to the T0 time point (plotted as purple points). A black line of gradient 1 is shown as a guide. Boxplots show the 25th, 50th and 75th quantiles, whiskers extend to the value at most 1.5 * IQR from the respective hinge.

(E) Likelihood functions for the age of individual GATC transcripts from the T0, T4, T8 and T16 time points of the HEK calibration are shown as curves, coloured by time point. The Maximum Likelihood Estimates are displayed, along with editing at 12 sites (intervening bases not shown).

(F) The MLEs are calculated for all GATC transcripts and presented as histograms with a non-parametric (LOESS method) smoothed line overlaid.

(G) MLE distributions are shown as violins for the T0 (navy) and T16 (yellow) calibration time points. The mean age is shown as white circles with the Δ age written above.

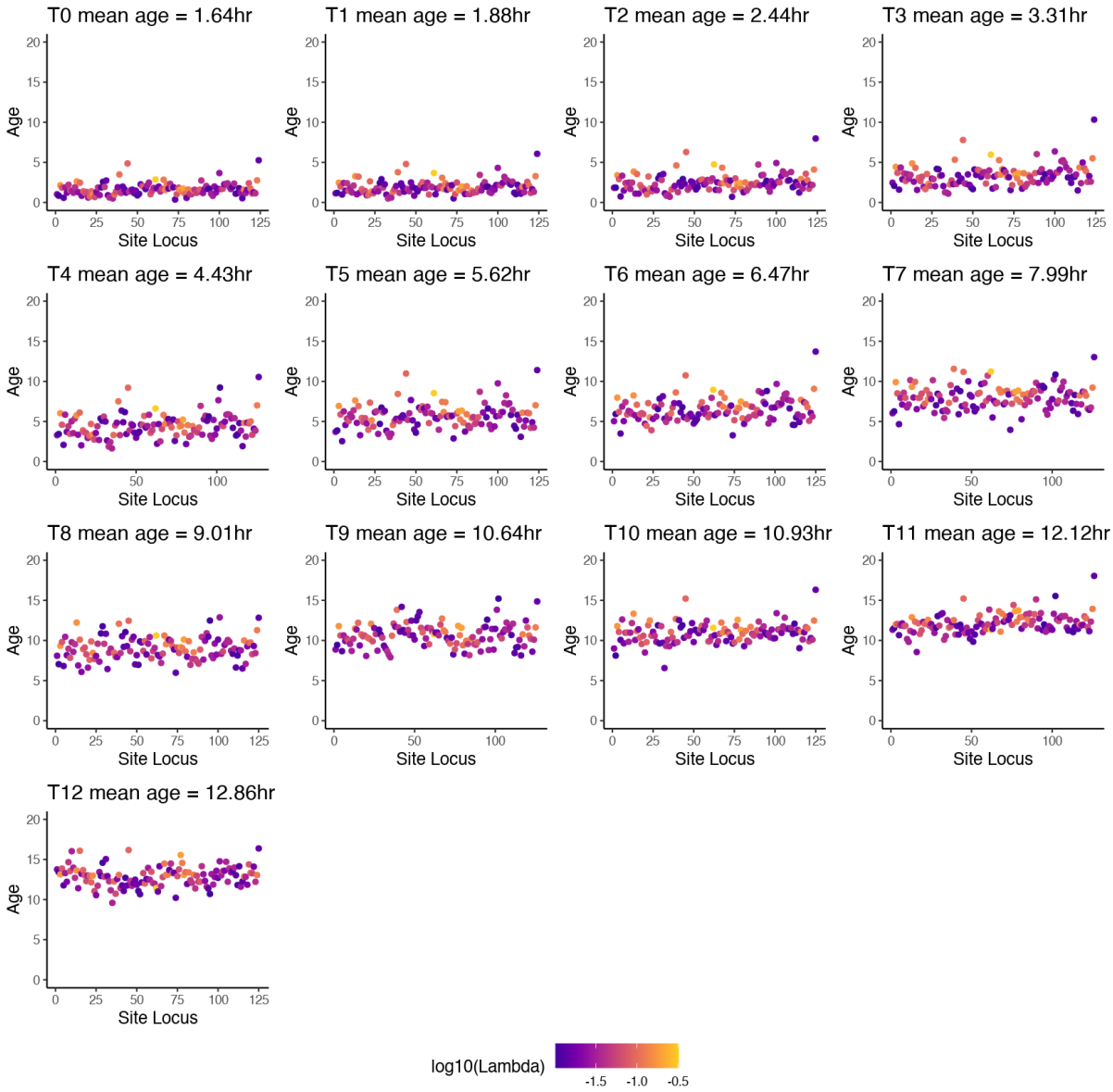


Figure 14 Predicted age form individual sites on the VHL gene in the HEK293 calibration experiment.

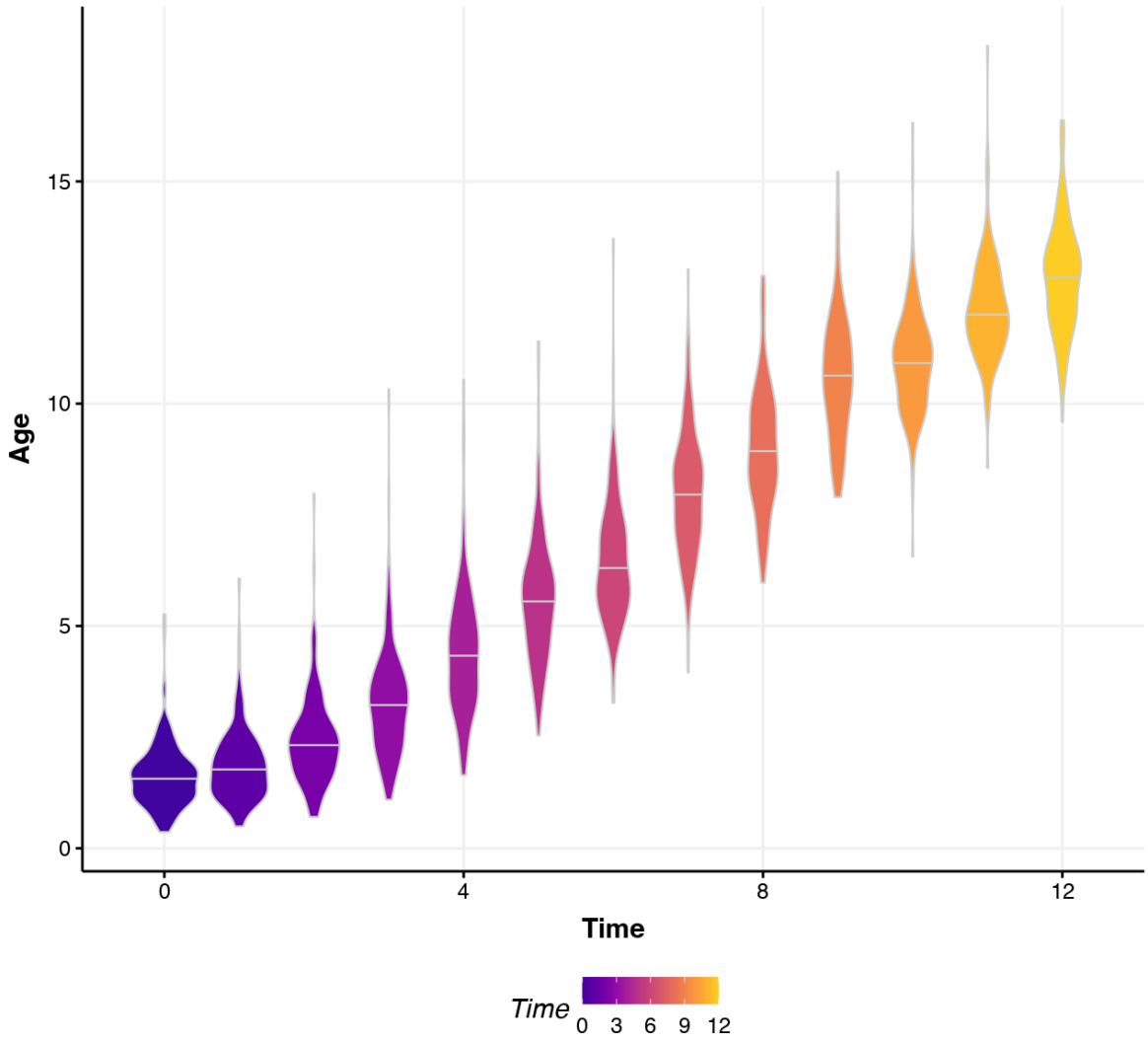


Figure 15 Mean gene age for VHL - calculate by taking the mean of the per-site ages shown in Figure 14.

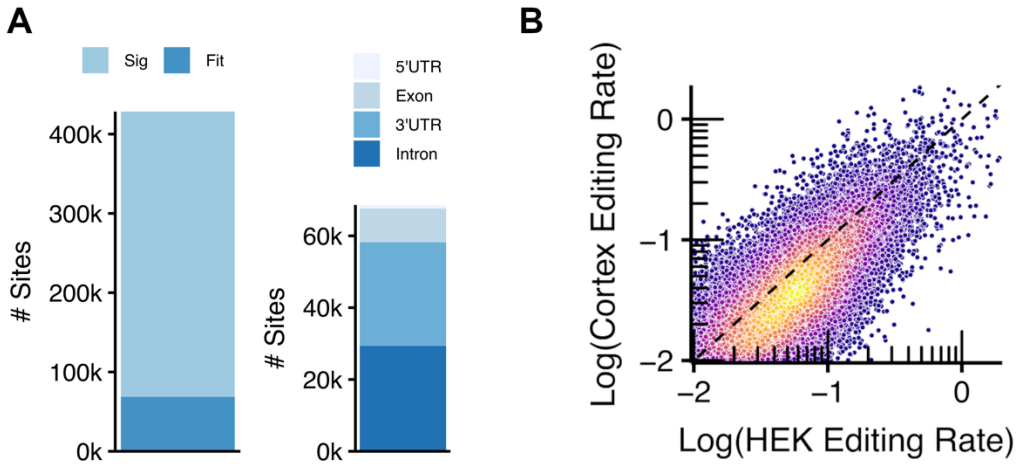


Figure 16 HEK293 calibration data and correlation of editing rates with cortex data.

(A) Left: 428,148 adenosines are significantly differentially edited in HEK cells expressing an ADAR2(E488Q)Nlambda plasmid 8 hours after adding actinomycin D. Right: Of those editing sites, 68,599 are well fit by an exponential CDF, and their location within genes are shown as stacked bars.

(B) The correlation of editing rates at shared sites in the ADAR2(E488Q)-Nlambda expressing HEK293 calibration and hiPSC-derived cortical neuron calibrations are shown as points coloured by 2D kernel density and plotted on log10-log10 axes.

10.2.7. Investigating the contribution of Nlambda-ADAR editing

While it seemed clear that human cells expressing a hyperactive, promiscuous editing construct could also be used to timestamp mRNAs, the additional performance benefit of Nlambda-ADAR was not clear. For example, after filtering, the distribution of editing sites across the transcriptome was very similar to the cortical neuron dataset (Figure 17A). Indeed, at editing sites which are shared between the cortical neurons and "hyper-edited" HEK293 cells, the per-site editing rates are highly correlated (Figure 17B).

To find out if Nlambda-ADAR was adding additional useful sites, I performed another calibration experiment, this time on HEK293 cells which were not expressing any construct, and therefore would only have endogenous levels of editing. I found that while Nlambda-ADAR edited throughout the transcript body, I still found that most useful sites were in the 3' UTR, indicating these might just be endogenous sites. Additionally, most sites are shared between Nlambda-ADAR HEK and endogenous HEK, and the portion of shared sites increases when upon filtering (Figure 18).

I next analysed sites by editing rate in all three of the calibration experiments before any filtering was applied (cortical, Nlamdba-ADAR HEK293 and endogenous HEK293). It now became clear that the additional sites conferred by Nlamdba-ADAR had very low editing rates and were largely not passing the editing rate filter ($\lambda \geq 0.01$) (Figure 19B). While the performance of Nlambda-ADAR was disappointing, I realised that non-neuronal cells had comparable levels of editing to neurons (Figure 19A and C), something which surprised us.

Nlambda-ADAR edits at both existing endogenous editing sites, and also many additional sites which are not seen in vivo. These additional sites are distributed throughout the transcript. While initially, Nlambda-ADAR appeared to be a success, on closer examination the additional editing sites conferredby Nlamdba-ADAR are edited at extremely low rates. The editing rates at existing sites appear to be only marginally increased, while the new sites are eddited too rarely to be useable.

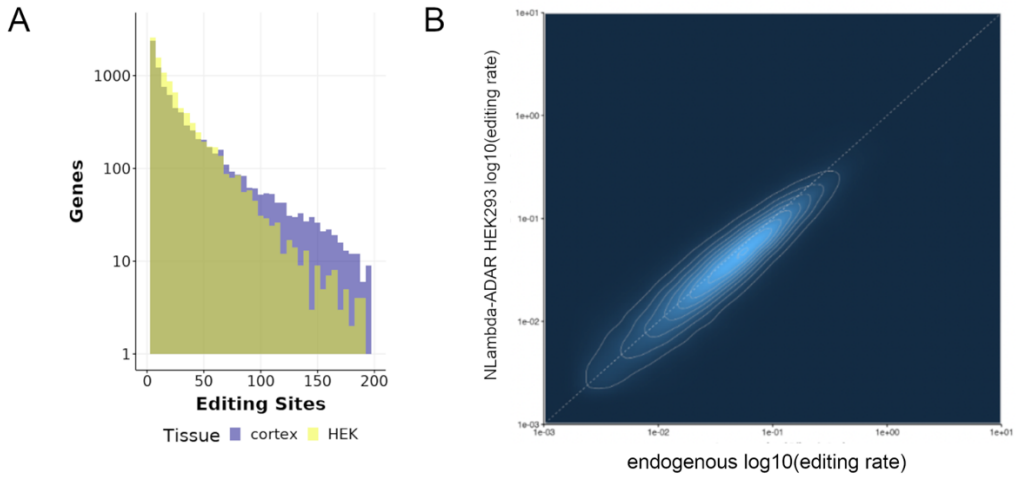


Figure 17 Nlambda-ADAR does not add many genes to Timestamps

(A) Distribution of editing sites across the transcriptome in the cortex and Nlamdba-ADAR HEK293 datasets.

(B) Shared sites in the Nlambda-ADAR and Cortex datasets have highly correlated editing rates (Pearson cor. = 0.88).

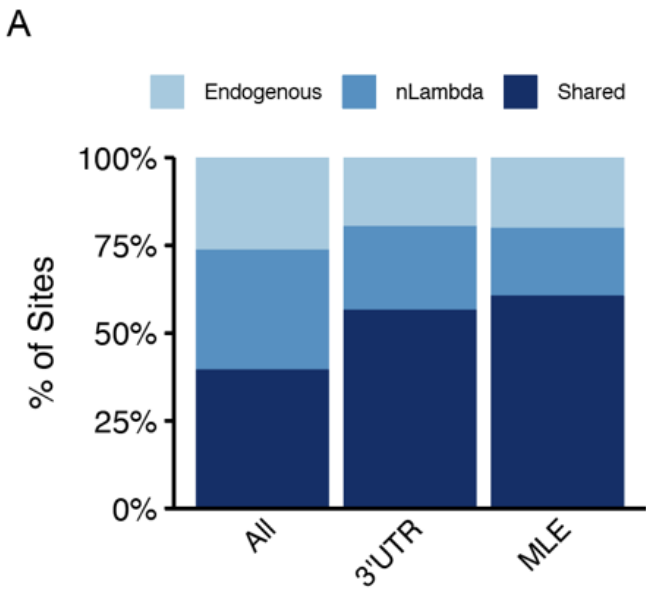


*Figure 18 Stacked bars showing the portion of sites from the Nlamdba-ADAR and endogenous HEK293 data which are endogenous only, Nlambda-ADAR only and shared. (**All** = all sites before filters; **3' UTR** = sites mapping to the 3'UTR only; **MLE** = sites which pass filters)*

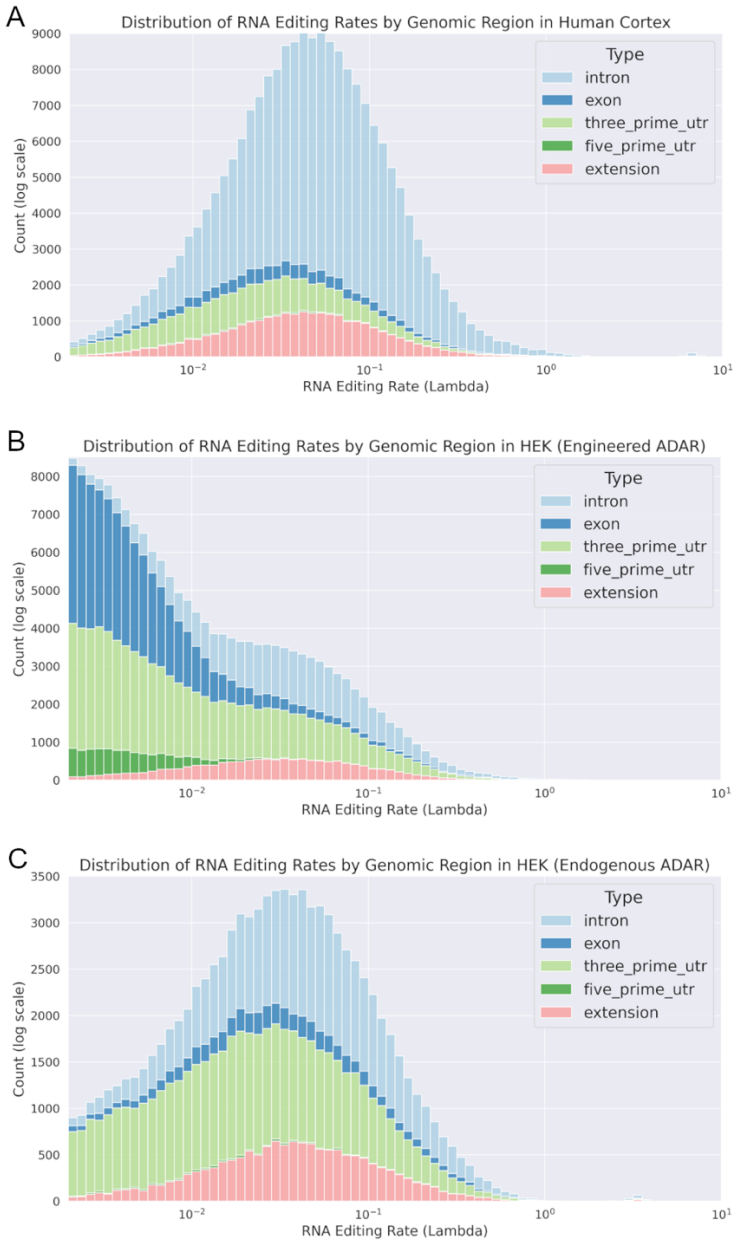


Figure 19 Genomic distribution of Nlambda-ADAR and endogenous A-to-I edits (analysis performed by rotation student David Miller).

(A) Histogram of per-site editing rates (λ from Equation 1) with stacked bars showing the genomic location (iPSC-derived human cortical neurons)

(B) Histogram of per-site editing rates (λ from Equation 1) with stacked bars showing the genomic location (Nlambda-ADAR HEK293 cells).

(C) Histogram of per-site editing rates (λ from Equation 1) with stacked bars showing the genomic location (un-transfected HEK293 cells).

10.3. Editing in other cell types

I cultured human brain slices, recently isolated from patients suffering with focal epilepsy and intestinal organoids, to study the potential for generalizing the timestamps method to other tissues. Despite substantial RNA degradation in both samples, I found editing rates to be highly correlated between these divergent cell types and the iPSC-derived cortical neurons Figure 20.

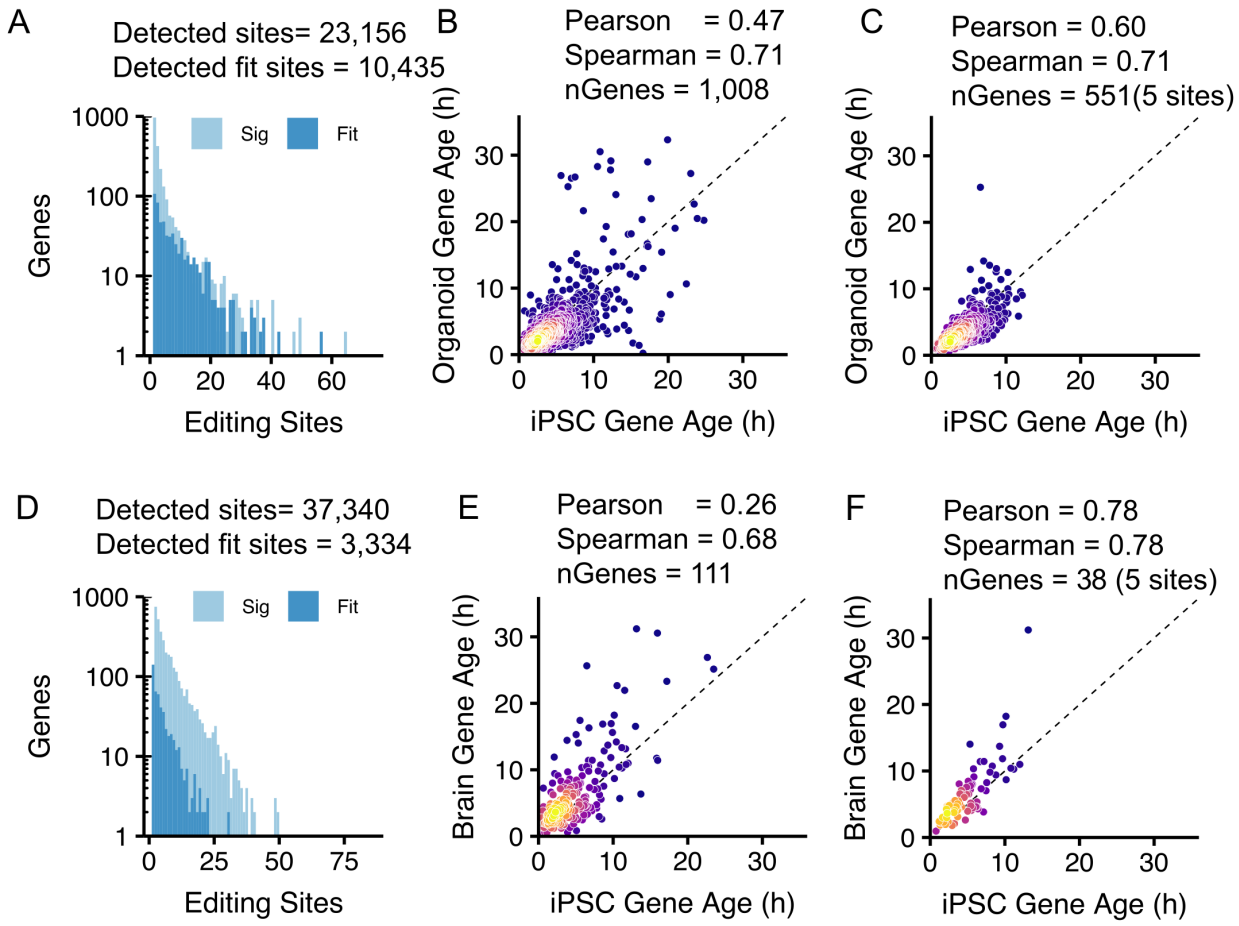


Figure 20 Editing sites in human intestinal organoids (A-C) and human brain slices (D-F)

(A and D) The distribution of editing sites across the transcriptome in human intestinal organoids and brain slices, respectively.

(B and E) Correlation of mean ages of genes with at least 1 fit site from the cortex calibration.

(C and F) Correlation of mean ages of genes with at least 5 fit sites from the cortex calibration.

10.3.1. Chapter Summary

In summary, Timestamps allows estimation of the ages of single molecules of RNA in divergent cell types (HEK293 cells and cortical neurons). I attempted to develop a hyperactive and promiscuous editor of RNA to boost the numbers of edits and broaden their distribution across the transcriptome so that more genes could be timestamped and at a higher temporal resolution, in addition to timestamping in species which do not have high levels of endogenous editing. This attempt fell short because the additional editing sites conferred by Nlambda-ADAR were infrequently edited.

However, to my surprise the endogenous editing sites in HEK293 cells were sufficient to timestamp thousands of genes. My belief that neurons would have much higher levels of editing than other cell types was an over estimate; one that others before us have made (Ramaswami et al. 2013; Paz-Yaacov et al. 2010; Pinto, Cohen, and Levanon 2014). As explained in the introduction, ADAR editing is in fact ubiquitous in human tissues (Tan et al. 2017). It is possible of course that there are cell types within these tissue with low levels of ADAR editing, but the ubiquitous distribution at the tissue level is enough to dissuade the idea that high levels of ADAR editing are only seen in neurons.

As a result of the high levels of editing which I observed in unmodified HEK293 cells, I decided to pursue studying the utility of Timestamps in these and other human cells, without expression of exogenous editing constructs. In parallel, I would continue to develop the second generation of hyperactive, promiscuous editors, as these could still have utility in Timestamping genes which are not naturally edited and extend Timestamps to organisms with low levels of editing (See section 13.4).

10.4. Methods for Chapter 10

10.4.1. HEK293

To explore editing in other cell types, and to obtain more timepoint and replicate data, I used a readily available laboratory cell line (HEK293). Unlike the iPSC-derived neurons, HEK293s do not require a lengthy differentiation period, are extremely robust to different perturbations and transfect readily, which would allow us to prototype hyperactive editors. Obviously, HEK293s would also come with the disadvantage of being profoundly aneuploid, potentially carrying my complex translocations and SNPs, and being a poorer model of human cellular physiology.

HEK293-FT cells were maintained in Dulbecco's Modified Eagle Medium (DMEM, Thermo Fisher, 11965092) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin-streptomycin (Sigma). For the experimental setup, HEK293T cells were plated in triplicate onto 24-well tissue culture plates 48 hours prior to the initiation of the experiment.

10.4.2. Human hiPSC derived neurons

Human induced pluripotent stem cells (hiPSCs) were derived from reprogrammed fibroblasts obtained from healthy donors, following ethical approval from the London-Hampstead Research Ethics Committee and the Joint Research Office of University College London, Great Ormond Street Institute of Child Health, and Great Ormond Street Hospital. The hiPSCs were cultured on Geltrex (Thermo Fisher) using E8 medium (Thermo Fisher) or mTeSR medium (Stem Cell Technologies) and were passaged with 0.5 mM ethylenediaminetetraacetic acid (Thermo Fisher). Neuronal differentiation was performed according to a previously established protocol (Shi, Kirwan, and Livesey 2012). In brief, neocortical stem cell differentiation was initiated through dual SMAD inhibition using SB431542 (10 μM, Tocris) and dorsomorphin dihydrochloride (1 μM, Tocris) for 12 days, followed by an extended culture period of 90 days to achieve neuronal maturation.

10.4.1. Human Organotypic Brain Slices

Organotypic brain slice cultures were prepared using the interface method adapted from Ravi et al. (Ravi et al. 2019) and De Simoni & Yu (De Simoni and MY Yu 2006). Fresh, healthy cortical human brain tissue was obtained during neurosurgical procedures, under National Research Ethics approval (Reference: 21/SC/0111). The resected tissue was immediately placed in ice-cold dissection medium in the operating theatre and promptly transferred for sectioning at a thickness of 300 µm using a Leica VT1200s vibratome. Each 1 cm^2 brain section was then plated onto a 30 mm culture plate insert (Millipore, PICM03050) within a 35 mm Nunc 6-well plate, which was filled with 1 ml of culture medium per well, and subsequently incubated at 37°C with 5% CO2. The culture medium was replaced every 24 hours. The dissection medium consisted of Hank's Balanced Salt Solution (HBSS) supplemented with HEPES (2.5 mM, pH 7.4), D-glucose (30 mM), CaCl2 (1 mM), MgCl2 (1 mM), and NaHCO3 (4 mM). The culture medium comprised Neurobasal l-Glutamine supplemented with 2% serum-free B-27, 2% Anti-Anti, D-glucose (13 mM), MgSO4 (1 mM), HEPES (15 mM), and GlutaMAX (2 mM). RNA extractions were conducted in triplicate using the Qiagen RNeasy Plus Universal Mini Kit, adhering to the manufacturer's protocols.

10.4.2. Organoid Culture

Human intestinal organoid cultures were established according to previously described protocols (Sato et al. 2011). In summary, organoids were revived from cryopreserved stocks, originally derived from human intestinal biopsies (Meran et al. 2020) (Research Ethics Committee references 04-Q0508-79 and 18/EE/0150), and maintained as three-dimensional spheroids in Cultrex™ reduced growth factor basement membrane extract, type 2 (R&D Systems). The organoids were cultured in 24-well plates using human IntestiCult Organoid Growth Medium (STEMCELL Technologies) supplemented with 3 µM CHIR99021 during the expansion phase, with a passaging ratio of 1:5 and media changes every 2 days. RNA extractions were carried out in triplicate using the Qiagen RNeasy Plus Mini Kit, following the manufacturer's instructions.

10.4.3. Calibration Protocol

Cells, either HEK293 or hiPSCs, were cultured according to the protocols previously detailed. Actinomycin D was added to the cultures at a final concentration of 1 µg/ml in complete media at time t = 0 hours. Cells were lysed in triplicate at predetermined time intervals. For example, HEK293 cells were lysed at the following time points: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 24 and 38 hours.

RNA extraction was performed using the RNEasy Plus Mini Kit (Qiagen, 74136) according to the manufacturer's protocol. Cell lysis was conducted directly in the culture wells using RLT Plus buffer from the Qiagen kit, followed by vortexing to ensure homogenization. The resulting purified total RNA was quantified using a Qubit fluorometer (Life Technologies, Q32855), and RNA quality was assessed using an Agilent Tapestation with RNA ScreenTape (Agilent, 5067-5576).

Full-length, short-read sequencing libraries for Illumina were prepared by poly-A selection using oligo-dT magnetic beads (NEB, E7490), followed by library preparation with the Ultra II Directional Library Prep Kit for Illumina (NEB, E7760), in accordance with the manufacturer's instructions. Briefly, RNA was fragmented, reverse transcribed with random primers, and adapters were ligated to both ends of the resulting cDNA. Strand specificity was ensured by incorporating uracil during second-strand synthesis, followed by digestion with USER enzyme. Sequencing adapters and sample barcodes were added via PCR, followed by purification, quality control, and pooling. Short-read sequencing was performed on an Illumina NovaSeq platform using 150-cycle kits, with paired-end reads of 76 base pairs each. The HEK293 samples were sequenced to a depth of approximately 225 million reads per time point, while cortical samples were sequenced to a depth of approximately 500 million reads per time point.

For long-read sequencing, both HEK293 and iPSC-derived cortical samples were analysed using PacBio Iso-Seq on a Sequel IIe platform. Library preparation was conducted with the

Iso-Seq Express Template Preparation Kit (PacBio) following the manufacturer's protocol. The cortical samples at both time points were sequenced on a single Sequel IIe SMRT Cell 8M flow cell in CCS mode, achieving a mean depth of approximately 2.2 million reads. In subsequent long-read sequencing experiments, the Oxford Nanopore Technologies platform was employed due to its higher throughput, which demonstrated comparable performance in determining single-molecule read lengths and quality.

10.4.4. PacBio versus Nanopore Sequencing

While PacBio sequencing produces long and highly accurate consensus reads, Nanopore produces long and relatively noisy reads even with recent improvements in pore architecture and chemistry (Bogaerts et al. 2024; Ni et al. 2023). Nevertheless, I believed nanopore would be critical for this project for two reasons: (i) the critical bases which need to be called with high accuracy are editing sites; the remainder of the sequencing data is simply for alignment to the reference genome and (ii) nanopore sequencing produces a much higher output than PacBio. For instance, in sequencing the 1-hour and 8-hour timepoints of the HEK293 data, I produced 90 million reads from an ONT PromethION flow cell versus 4 million on a PacBio Sequel II flow cell. Figure 21 Nanopore per-base quality scores (QVs), shows that while the (per-read) modal quality scores reported by ONT's MinKNOW software were relatively poor, much higher modal (per-base) QVs are seen when filtered to editing sites only. This can be explained by the fact that clusters of low-quality base calls (mostly in the first and last 100 bases) drag down the average quality of each, otherwise accurate, nanopore read (Delahaye and Nicolas 2021). Therefore, after the initial long-read sequencing of the HEK293 and cortex calibration experiments, I decided to switch from PacBio to ONT in subsequent chapters.

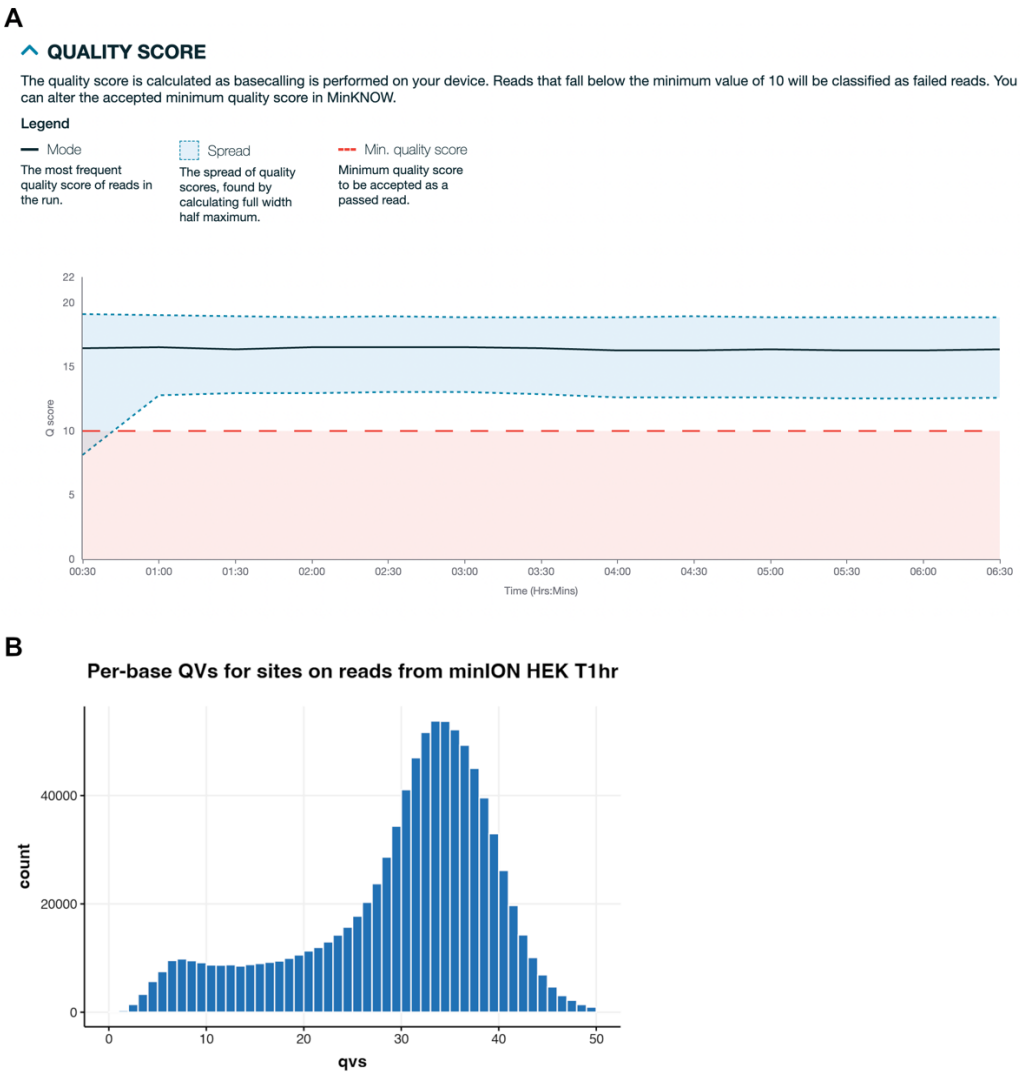


Figure 21 Nanopore per-base quality scores (QVs)

(A) A screenshot from Oxford Nanopore MinKNOW software showing the pre-alignment per-read QVs calculated based on uncertainty in Nanopore's recurrent neural networks.

(B) Histogram showing post-alignment (Minimap2) per-base QVs at editing sites only (mode = 34).

10.4.5. Hyperactive editor experiments

HEK293 cells were transfected with engineered, hyperactive RNA editors (and corresponding guide RNAs if required). Twenty-four hours later, I added actinomycin D to

stop transcription and lysed the transfected HEK293 cells at 0-hour and 8-hour timepoints in triplicate.

In the first generation of editor selection (this chapter), I studied the full-length hyperactive ADAR2 mutant ADAR2(E488Q). Additionally, editors were fused to the ADAR2(E488Q) catalytic domain and tested in this round:

- **Nlambda-ADAR2(E488Q)** (Montiel-Gonźalez, Vallecillo-Viejo, and Rosenthal 2016),
- **SNAP-ADAR2(E488Q)** (Vogel et al. 2018),
- **Cas13-ADAR2(E488Q)** (Cox et al. 2017),
- **PABP-ADAR2(E488Q)** (courtesy of Rory Maizels at the Francis Crick Institute).

I also tested ABEmax and the cytosine base editor BE3 (Grünewald, Zhou, Iyer, et al. 2019). Although SNAP-ADAR2(E488Q) showed slightly higher levels of editing than ADAR2(E488Q)-Nlambda, I decided not to proceed with it because its guide RNA cannot be genetically encoded due to the need for an O6-benzylguanine base.

All plasmids used in this study are listed in Table 8.

10.4.6. Bioinformatics

R4.1.0 (https://www.r-project.org/) was used to undertake all analysis unless otherwise specified.

10.4.6.1. Short-read data processing

Bcl files generated from short-read sequencing (Illumina technologies) were converted to FASTQ format using bcl2fastq2 version 2.20.0 (https://emea.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html). The resulting FASTQ files were then trimmed using cutadapt version 3.5 (https://github.com/marcelm/cutadapt/) and subsequently aligned to the GRCh38.100

human reference genome with HISAT2 version 2.1.0 (http://daehwankimlab.github.io/hisat2/). Gene-level quantification of read counts was performed using salmon version 1.4.0 (https://combine-lab.github.io/salmon/). RNA editing events were quantified using JACUSA2 version 2.0.4 (https://github.com/dieterich-lab/JACUSA2) with the call-1 (detect) mode, applying filters to exclude potential editing sites located near the beginning and end of reads, indels, splice sites, and within homopolymer runs exceeding 7 bases. The output files from JACUSA2 were converted into BED format using a custom Python3 script.

10.4.6.2. Long-read data processing

POD5 files obtained from ONT sequencing were basecalled using Guppy version 6.4.6-CUDA-11.7.0, with the Super Accuracy Mode (SUP) enabled, on NVIDIA A100 and V100 GPUs. The generated FASTQ files were then demultiplexed using the guppy_barcoder tool from the same Guppy version.

The demultiplexed FASTQ files, whether from ONT or PacBio, were processed through a unified analysis pipeline. Reads were aligned to the GRCh38.100 human reference genome using minimap2 (https://github.com/lh3/minimap2) and subsequently quantified with IsoQuant version 3.0.0 (https://github.com/ablab/IsoQuant). RNA editing events were quantified following the same method used for short-read data as described earlier.

For estimating the age of single molecules, base calls at specified loci were extracted directly from the .bam files using custom R scripts, specifically the script named `extract_calls.R`, which is available in the GitHub repository associated with this project. Long-read data from individual cells were deduplicated using the `markdup` function from samtools version 1.13 (https://www.htslib.org).

10.4.7. Modelling and Statistics

10.4.7.1. Editing rate determination from calibration data

Sites exhibiting significant increases in RNA editing at 8 hours following the addition of ActD were identified using JACUSA2 call-2, applying a filter to retain sites with a Z score greater than 1.96 (termed "Significant sites"). A linear regression model was then fitted to each site, and sites with a non-positive gradient were excluded, as these indicated either no accumulation of edits or the presence of a potential single nucleotide variant (SNV).
Given that the proportion of reads edited at each site is likely non-zero at the initial time point (0h), the standard editing rate equation:

$$y_i(t) = 1 - \mathrm{e}^{-\lambda_i \mathrm{t}}$$

was adjusted by introducing an additional parameter, α, resulting in the calibration equation:

$$y_i(t) = 1 - \alpha_i \mathrm{e}^{-\lambda_i \mathrm{t}}$$

Equation 1

In this equation, y_i represents the fraction of edited reads at site i, α is a free parameter estimating the edited fraction at the start of the calibration experiment, λ is the editing rate, and t is the time elapsed since the experiment's onset.

The modified calibration equation was fitted to each remaining site using non-linear least squares regression, implemented via the `nls` function from the R package `stats` v4.1.2, with initial values set to λ=0.01 and $a = 1 - y(t = 0)$. The list of editing sites was then refined by applying the following criteria to define "fit sites":

- $R^2 \geq 0.4$,

- $2 \geq \lambda \geq 0.01$
- Each site must be present in at least 4 time points within the calibration dataset.
- The predicted age of the site at the T0 calibration time point should be less than 25 hours.

Sites were annotated as belonging to a specific gene if their genomic coordinates fell within the region spanning from the start of the 5'UTR to the end of the 3'UTR and matched the sense strand. Sites located in regions annotated with multiple genes on the same strand, where gene assignment could not be clearly determined, were excluded from the analysis. Additionally, sites mapping to intergenic regions were also removed from the list of fit sites. Calibration datasets from hiPSC-derived cortical neurons and HEK293 cells (transfected with ADAR2(E488Q)-Nlambda) were compared by calculating the base-10 logarithm of the editing rates and determining the Pearson correlation coefficient for the shared sites. Given that the editing levels in HEK293 cells transfected with ADAR2-Nlambda were found to be like those in un-transfected HEK293 cells, the ADAR2-Nlambda calibration data were used for all experiments in this study, as it encompassed more time points. The datasets were integrated by calculating the mean log fold change at all shared sites, adjusting the editing rates at hiPSC-only sites by this factor, and appending them to the list of HEK293 editing sites. The final set of 201,233 fit sites used for analysis was derived from this combined dataset as described above.

10.4.7.2. Per-site age

To estimate the mean age (t) of the transcript population at a specific site i, the editing rate equation was rearranged. By using a known editing rate (λ) and the observed proportion of edited reads (y), the equation was rearranged as follows:

$$t = -\frac{\ln(1 - y_i)}{\lambda_i}$$

Equation 2

To determine the mean age of a specific gene, the average of the age estimates for all annotated sites within that gene was calculated.

To explore how the mean age of each gene changes throughout the calibration experiment, I computed the average site age at each time point for each gene (n = 5,540) and then subtracted the age at T0. Genes that lacked sites at the T0 timepoint were excluded from the analysis. To prevent the issue of sites with infinite age, any site that was fully edited across all reads was adjusted by subtracting 1 divided by the total number of reads before calculating the age.

10.4.7.3. Single molecule age

For the long-read sequencing data, base calls at fit sites were extracted from the .bam files. I then construct a likelihood model for the age of a transcript ("single molecule age") by modelling each fit site as a sample from a time-dependent binomial distribution. This function estimates the transcript's age based on the observed editing state, x, across its sites:

$$\mathcal{L}(x,t) = \prod_{i \in G}(1 - e^{-\lambda_i t}) \prod_{j \in A} e^{-\lambda_j t}$$

Equation 3

Here, set G includes sites that are edited, while set A contains unedited sites.

$$L(t;x) = \sum_{x \in X} \frac{d(t;x)}{\int_0^\infty d(t;x).\,dt}$$

Equation 4 the per-gene density function is the sum of the normalised per-transcript likelihood functions.

When estimating the age of single molecules, I considered only genes with at least 5 fit sites within the 3' UTR, resulting in a dataset of 32,803 sites across 949 genes. The mean age of the reads was determined through maximum likelihood estimation (MLE), using the `maxLik` function (R package `maxLik` v1.5-2) to maximize the log-likelihood function for each transcript.

To estimate the population distribution for any gene, I generated a histogram of the MLE values (typically ranging between 0 and 20 hours, with a bin width of 2 hours). Depending on the analysis, I either used raw counts of MLEs per bin or normalized the counts to obtain a density estimate. When a smoothed line was required, it was generated using the LOESS method within the `ggplot2` package (v3.4.1). For visualizing connected points with error bars, the points represent the mean values across replicates, with error bars indicating the standard error of the mean. Mean gene ages are the mean of the individual transcript MLEs.

11. RNA editing records the cell's recent transcriptional history.

Initial experiments showed that Timestamps was accurately measuring the ages of single mRNA molecules when measured against ground truth (the intervals between timepoints in the calibration experiments). I next sought to determine if past transcriptional activity could be uncovered from the ages of single molecules of mRNA. One can imagine that MLE distributions such as that in Figure 13F could, if long-read sequencing was performed on a cell freely transcribing mRNA (not treated with ActD), inform the past transcriptional history of the cell. This is analogous to the use of carbon dating in palaeontology, where it is used to determine the age of fossils. The age of individual fossils can then be used to create a narrative of the fossil record. (The analogy goes deeper, as calibration curves – records of the levels of Carbon-14 in the earth's atmosphere over the past 50,000 years are used to inform carbon dating models (Aitken 1990). Furthermore, carbon dating is typically limited to relatively recent fossils, typically those up to around 50,000 years old. Beyond this age, the remaining carbon-14 becomes too minimal to detect accurately. I may have a similar limitation in determining the cell's distant transcriptional history due to the short half-lives of mammalian mRNAs; typically 4 hours, although this differs by an order of magnitude between genes (Herzog et al. 2017).

11.1. Human Gene Induction Experiment

To test this strategy, I cloned full-length cDNA sequences (including exons and 3' UTR) of three genes identified from the calibration data (*ACBD7*, *ATG14* and *BVES*) into TET-ON expression plasmids. These genes were the first three alphabetically on a list of genes that (i) have >10 fit sites in the 3' UTR, and (ii) for which the total cDNA length was <5000 nucleotides. (These two properties were chosen to give a list of genes with precise age determinations, and which would be feasible to clone into a plasmid, respectively). My strategy for cloning these plasmids is outlined in 11.2.2.

HEK293 cells were transfected with an equimolar mixture of the three tetracycline-sensitive plasmids and then stimulated to produce either “ON,” “PULSE,” or “OFF” behaviours. I stimulated cells to produce these three transcriptional behaviours by using doxycycline and actinomycin D to switch transcription on and off, respectively (see section 11.2). Wells were lysed at several timepoints, and mRNA was prepared for long-read sequencing (Figure 22).

I hypothesised that:

- in the ON condition, there should be a rapid increase in the number of *ACBD7*, *ATG14* and *BVES* transcripts as transcription on the plasmids are switched on. This should have the effect of immediately reducing the mean age of all three genes (as measured by taking the mean of the MLEs).
- in the OFF condition, HEK293 cells were pre-treated for doxycycline so that plasmid transcripts were highly expressed, before adding actinomycin D to halt transcription. The number of transcripts in the cells was expected to decay as mRNAs were degraded. I expected the mean age therefore to increase in the OFF condition, as no new transcripts would be produced during observation.
- in the PULSE condition, I expected a peak in the MLE histogram, moving from younger to older ages.

Timestamps should be able to differentiate between all three behaviours.

To test the kinetics and leakiness of the TET-ON induction system, I performed a pilot experiment using only the ON condition so that the experiment could be sequenced on the lab’s minION (Oxford Nanopore Technologies). Snapshots from IGV (Figure 23) show that as time form doxycycline induction increases, new mRNAs are transcribed, and these new mRNAs have little editing. Histograms of the MLEs for each of the three induced genes show that new transcripts enter the system over the 8-hour induction, slowly from 0-4 hours and more rapidly thereafter. Despite the low sequencing depth from the minION, (the counts for ACBD7 are less than 30 at their peak at 8 hours), Timestamps clearly shows the production of new transcripts as early as 4 hours post-induction. For ATG14, which has the most counts,

one can clearly see the transcripts produced at the beginning of the experiment have aged and increased the counts in the older bins (Figure 24 – ATG14). The downward slope of the histograms is expected because of mRNA turnover, which is continuously removing transcripts from the system (for example, c.f. figure 4 parts a and b in (Herzog et al. 2017)).

The mean age did not decrease decisively after induction, which is not surprising considering the skew of the histograms towards younger ages, which makes the mean a poor measure of the central tendency as it is sensitive to transcripts at the old ages (the tail). Additionally, the sequencing depth was low on this (pilot) experiment, which meant the mean could be influenced by outlier MLEs. This is especially obvious if one compares the 0H to the 8H mock induction (8H_c) conditions. While the histograms look similar (as they should), the mean ages are very different (Figure 24).

I next performed the full experiment, with all here conditions: ON, OFF and PULSE (Figure 25). With a high sequencing depth from a promethION flow cell, the transcript counts were much higher. Although the plasmids were transfected at equimolar ratios, far more transcripts were produced by the ATG14 plasmid, probably a result of more efficient transcription of this cDNA. The histograms of all three conditions were as expected. However, both the ON and OFF conditions showed increasing mean ages with time from induction. This is counterintuitive for the ON condition but is likely because the transcript counts are very low at the beginning (before doxycycline addition) which means there are few old transcripts. Switching on transcription from the plasmids increases the number of older (transcripts and skews the mean to older ages. This means that mean age will not be useful for distinguishing between ON and OFF transcriptional activity *in vivo,* however they can be easily distinguished by the MLE histograms which move in opposite directions (Figure 25).

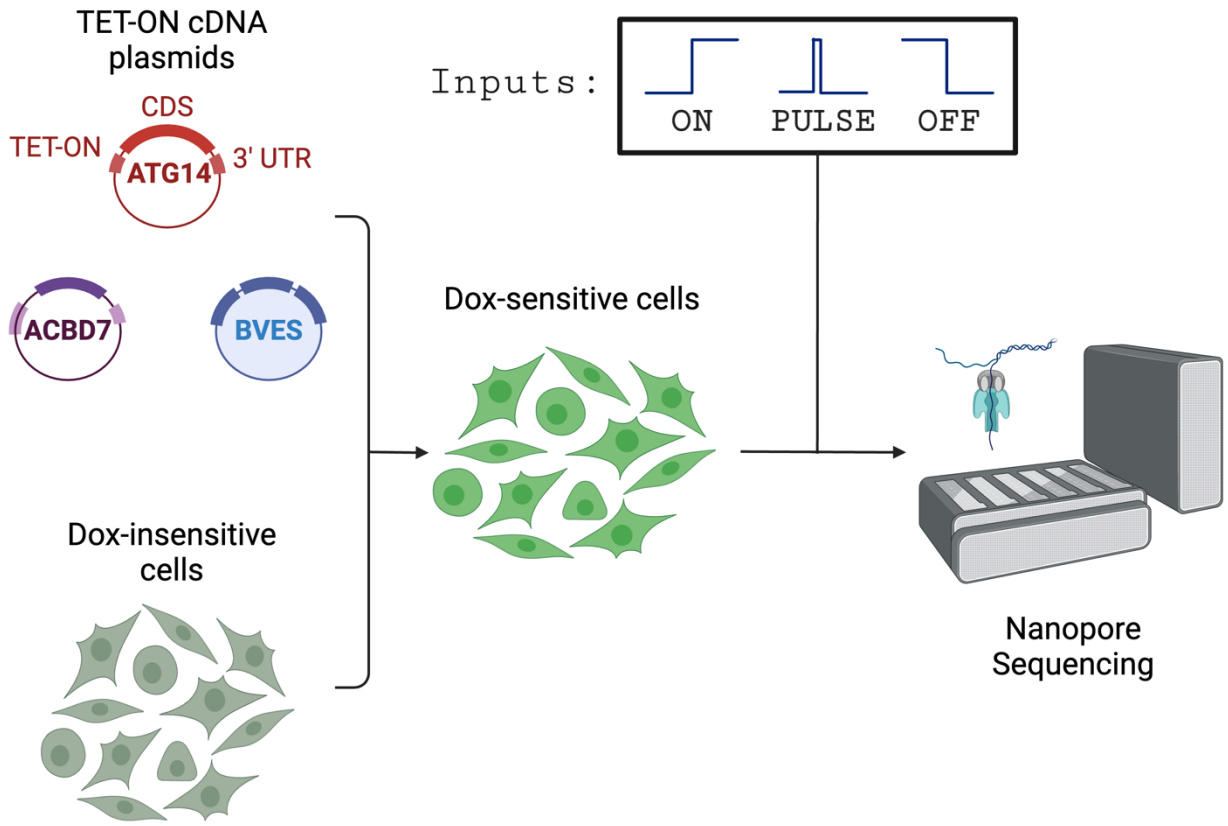


Figure 22 Schematic of the induction experiments carries out in HEK293 cells (produced using Biorender).

HEK293 cells were transfected with a mixture of three plasmids expressing the full-length human cDNAs for the genes *ATG14*, *ACBD7* and *BVES* under the control of the TET-ON system.

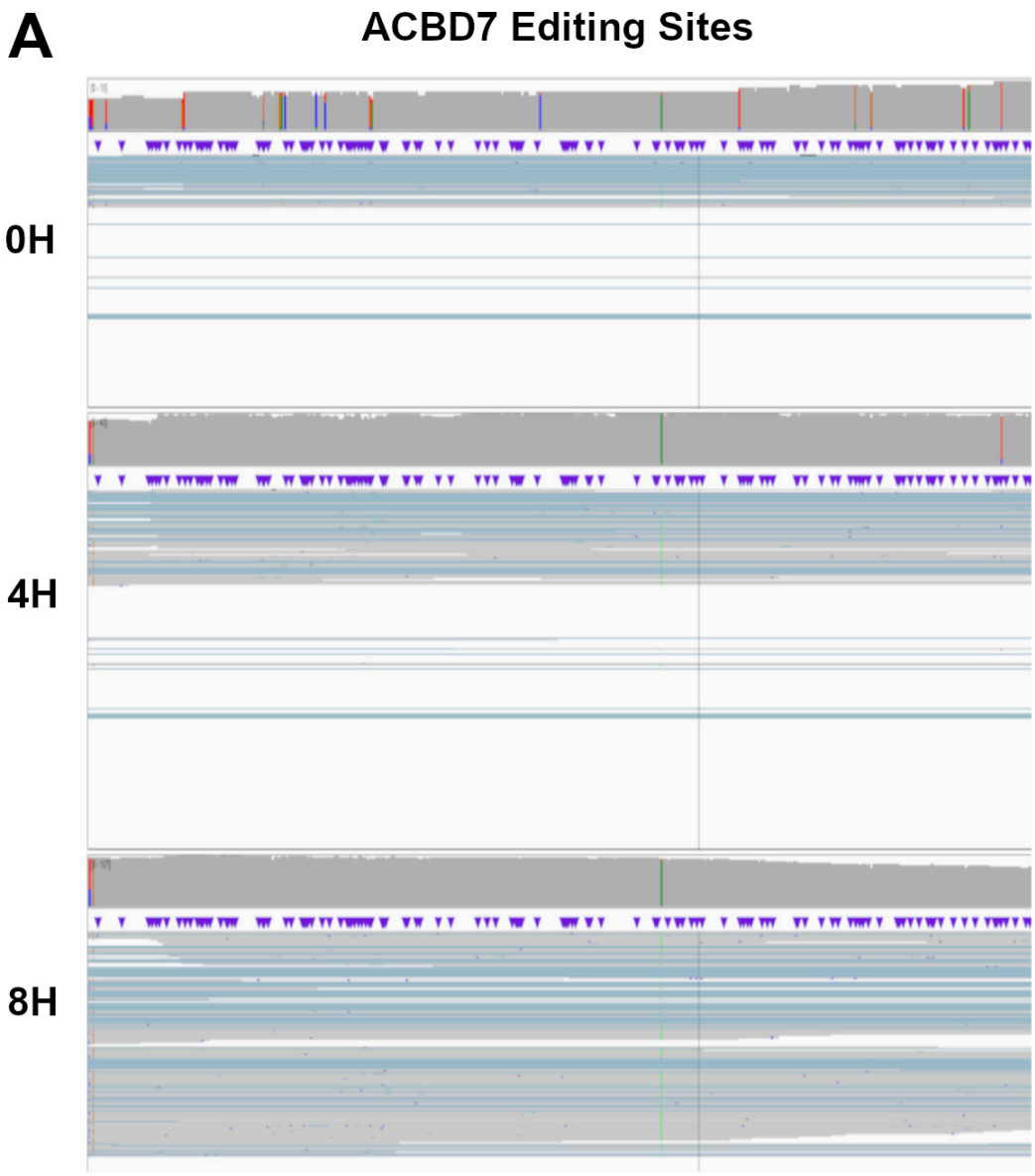


Figure 23 IGV screenshot of raw editing sites (pre-filtering) from three timepoints in the HEK293 induction experiment after doxycycline addition (ON condition; pilot experiment).

The screenshot from IGV genome browser shows the alignment of BAM files to the human reference genome for one replicate, for illustration purposes. BAM files encode the alignment of sequencing reads to a reference, describing for example the positions of mismatches or deletions. In the snapshots, the light blue horizontal lines are nanopore reads. The grey histogram at the top of each screenshot shows the normalized depth at each site in the genome (number of reads mapping). Red and blue vertical lines mark the positions of A-to-G transitions (A-to-I edits). The green vertical line is an SNP. IGV is set to show edits with at least 10% of total reads mapping to a locus.

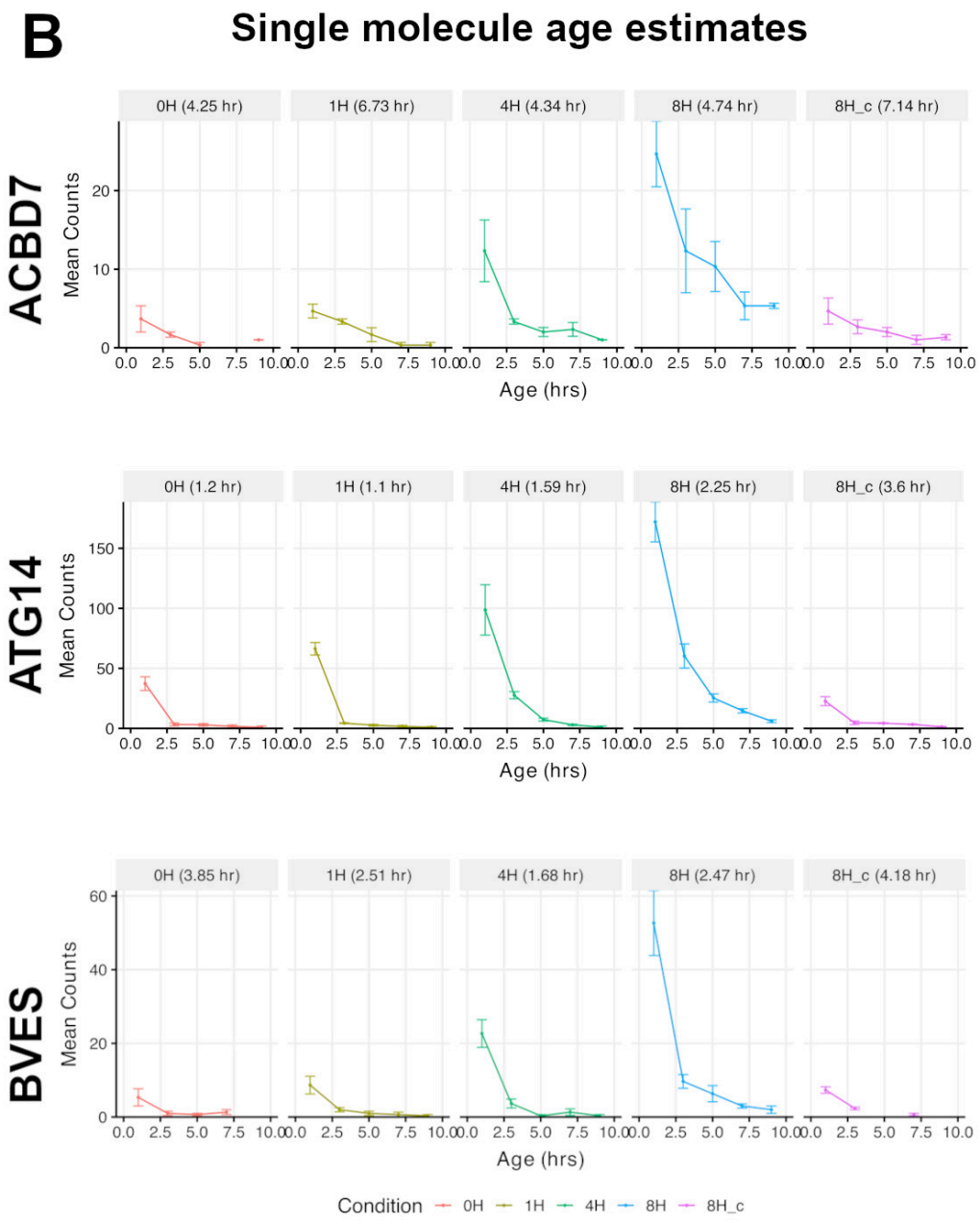


Figure 24 Histograms of transcript MLEs after doxycycline addition (ON condition; pilot experiment)

MLE histograms for ACBD7, ATG14 and BVES generated from long-read sequencing (n=3; bins=2.5hours; 95% CIs shown). The experimental timepoint is shown above each histogram with the mean age from the MLEs in brackets. (8H_c = 8-hour control – no doxycycline added at 0H).

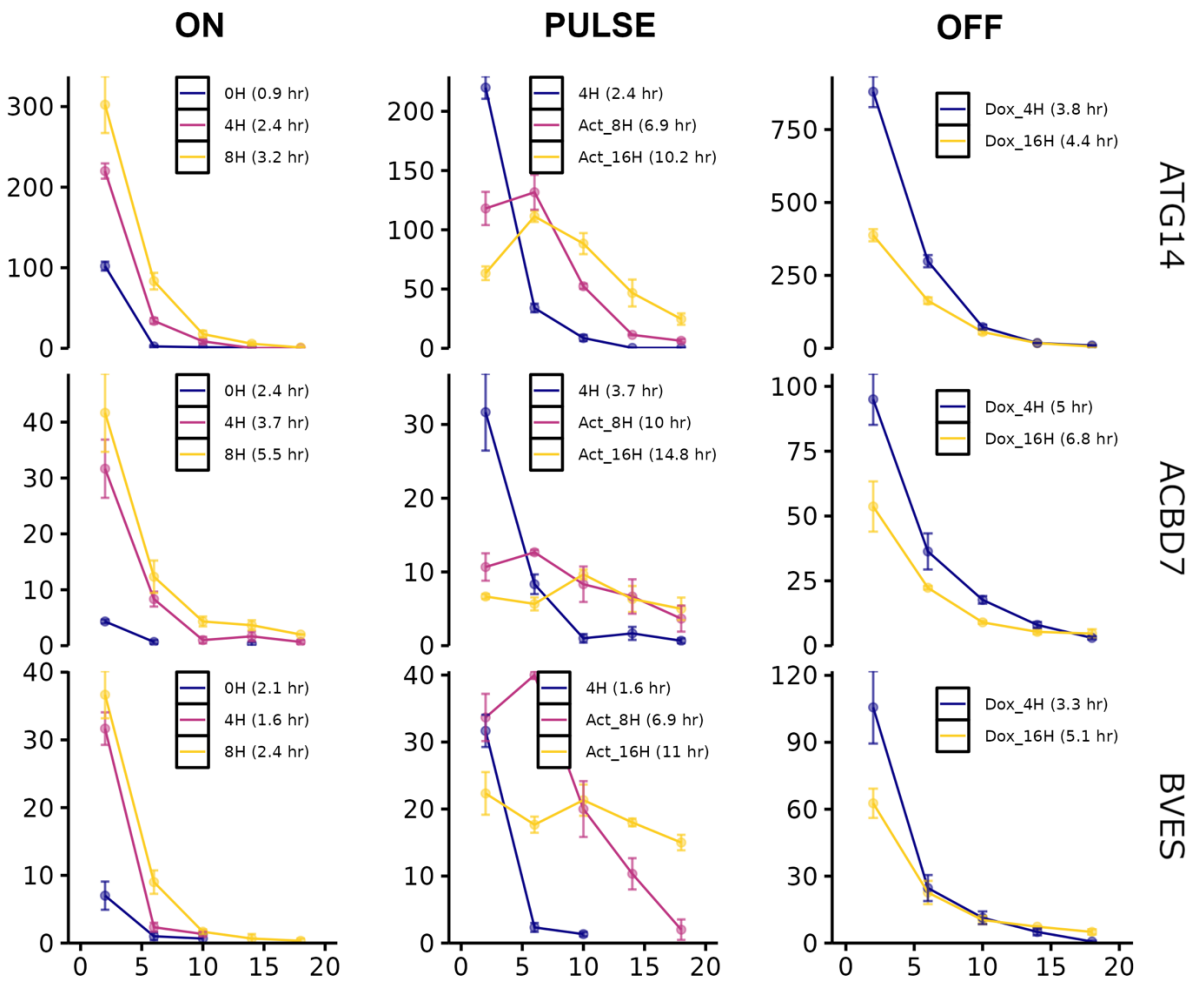


Figure 25 Matrix of MLE histograms showing the results of the ON/PULSE/OFF conditions for all three genes studied at all timepoints sampled. X-axis shows age in hours and Y-axis shows counts (n=3; 95% confidence intervals are shown).

11.2. Methods for Chapter 11

11.2.1. HEK293 cell culture

HEK293-FT cells were maintained in Dulbecco's Modified Eagle Medium (DMEM, Thermo Fisher, 11965092) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin-streptomycin (Sigma). For the experimental setup, HEK293T cells were plated in triplicate onto 24-well tissue culture plates 48 hours prior to the initiation of the experiment.

11.2.2. Cloning

Total RNA was extracted from HEK293 cells using the RNEasy Mini Kit (Qiagen), and a complementary DNA (cDNA) library was subsequently generated. The cDNAs for ATG14, BVES, and ACBD7 were amplified from this library using gene-specific primers. The resulting PCR products were gel-purified and assembled into a plasmid vector using NEBuilder Hi-Fi DNA Assembly (NEB, E2621). The assembled plasmids were then used to transform E. cloni 10G Chemically Competent Cells (Lucigen), which were cultured overnight.

Sequencing of the plasmid DNA enabled the identification of pre-edited ADAR editing sites within the sequences. Site-specific PCR was employed to revert these editing sites to their original, unedited forms. The resulting PCR products were reassembled using NEBuilder Hi-Fi DNA Assembly, and the accuracy of the final constructs was verified through whole plasmid sequencing.

The final constructs consisted of plasmids containing the full-length cDNAs for ATG14, BVES, and ACBD7, and were designated as pAG051, pAG053, and pAG054, respectively (Table 8).

11.2.3. Cell transfection

HEK293 cells were transfected using the TransIT-X2 Dynamic Delivery System (Mirus Bio) when the cells reached approximately 70% confluency. For each well of a 24-well plate, 250 ng of the tTA3 transactivator plasmid and 250 ng of an equimolar mixture of the three mammalian cDNA plasmids (pAG051, pAG053, and pAG054) were used for transfection. The transfection process was conducted 48 hours prior to the t=0 timepoint.

Three experimental conditions, designated as ON, OFF, and PULSE, were applied as follows:

- **ON Condition**: At t=0 hours, doxycycline was added to a final concentration of 1 µg/ml. Wells were lysed at 0, 0.5, 1, 4, and 8 hours. Additionally, a mock induction was lysed at the 8-hour timepoint.
- **OFF Condition**: Doxycycline was added 24 hours prior to the t=0 timepoint. At t=0 hours, the cells were transferred to doxycycline-free media, and lysates were collected at t=16 hours.
- **PULSE Condition**: Doxycycline was added at t=0 hours to a final concentration of 1 µg/ml. ActD was added at t=4 hours. Cells were lysed at t=8 and 16 hours.

For each condition, wells were lysed in triplicate at the specified timepoints.

11.2.4. Library preparation and sequencing

Long-read sequencing was performed on the Oxford Nanopore platform (see discussion on Nanopore versus PacBio in section 10.4.4). Library preparation was conducted using a customized SMART-seq protocol specifically designed to minimize the number of PCR cycles, thereby reducing the potential for PCR-induced errors (protocol courtesy of Adam Cribbs and Danson Loi). Briefly, a minimum of 100 ng of total RNA was denatured at 65°C in the presence of 100 µM poly-T reverse primer and 1 mM dNTPs. Reverse transcription was carried out by adding RT buffer (Thermo Fisher), Maxima H- reverse transcriptase (Thermo Fisher), and a template-switching oligonucleotide (Table 9 - Appendix). A custom

temperature ramp-up and cycling program was employed (Table 2). To remove excess primers, the reaction mixture was treated with heat-labile Exonuclease I (Thermo Fisher). Sample clean-up was then performed using SPRI-select beads (0.6X by volume).

The initial PCR amplification was conducted using 2X KAPA ReadyMix (Roche) and an ONT PCR handle primer. This was followed by a second round of PCR, during which the reaction was split into four separate reactions (details provided in Supplementary Table S6).
For sequencing, the samples were prepared using the Ligation Sequencing Kit v.14 (ONT, SQK-LSK-114) following the manufacturer's instructions. Sequencing was performed on a PromethION24 sequencer (ONT) across five R10.4.1 flow cells, yielding an average data output of approximately 70 Gb per flow cell.

Table 2 Thermocycling conditions for reverse transcription and PCR

Ramp-up + Cycle RT Programme:		
Temperature (ºC)	duration	cycles
8	12s	10x
15	45s	
20	45s	
30	30s	
42	2min	
50	3min	
50	20min	1x
4	Hold	1x
SMART PCR		
Temperature (ºC)	duration	cycles
95	3min	1x
98	20s	4x
65	45s	

72	3min	
98	20s	6x
67	20s	
72	3min	
72	5min	1x

11.2.5. Modelling and Statistics

For each condition, the transcript age was determined by averaging the MLEs. Due to the presence of short sequences aligning to the induced genes, I increased the threshold to a minimum of 10 sites per transcript (from the usual 5) to filter out these sequences. Transcript abundance was quantified using IsoQuant v3.0.0, and for easier interpretation, transcript per thousand (TPT) values were derived by dividing the transcript per million (TPM) values by 1,000. Clustering and heatmap visualization were performed using the `pheatmap` package (v1.0.12).

12. Differential mRNA age in activated human primary monocytes

12.1. Differential abundance versus differential age

My work so far has shown that Timestamps can measure the mean age of a gene (the mean age of an ensemble of transcripts transcribed from a single locus) and the age of single transcripts of mRNA.

Traditionally, changes in RNA abundance between experimental conditions and controls are used to study transcriptional regulation, utilising tools such as DESeq2 (Love, Huber, and Anders 2014). However, there are several reasons to believe that certain mechanisms of transcriptional regulation are missed by traditional differential abundance analysis. For example, small changes in transcription may cause only small relative changes to the total transcript abundance, especially if the transcript has a long half-life and so there is a large pool of transcripts in the cell. On the other hand, bursts of transcription could cause large changes in the mean age as only new transcripts can be produced by transcription. Additionally, if transcription is upregulated while transcript degradation is increased, transcript abundance would remain stable, while transcript age would decrease. Even more importantly, because transcriptional regulation is a dynamic process, interval sampling can miss transient yet critical changes in RNA expression. For example, closed chromatin can be rendered accessible through the transient expression of pioneer transcription factors (TFs) (Yu and Buck 2020), PU.1 in human monocytes is such an example (Zaret and Carroll 2011). This phenomenon may be extremely prevalent; across Eukarya, only a small fraction of TF-bound genomic targets identified by CHIP-seq can be validated in perturbation experiments. One possible explanation for this is many TFs may have only a transient, facilitatory role in transcriptional regulation (Swift and Coruzzi 2017).

Monocytes migrate into tissues and rapidly differentiate into tissue-specific macrophages in response to infection. They undergo global changes in gene expression required to adapt to

infectious challenges. LPS induced differentiation of primary human monocytes could provide an excellent model to test the capabilities of Timestamps in human tissues.

Primary human monocytes freshly isolated from human whole blood were stimulated with LPS and lysed after time intervals ranging from 0 to 6 hours (Figure 26). While the differential expression and differential aging analysis was performed on the 0H and 6H timepoints, the intervening timepoints allowed me to verify predictions of past transcriptional history produce from Timestamps. Differentially expressed genes between T0 and T6 were enriched for NFκB, TNF-α and NOD-like signalling pathway members, consistent with differentiation to an 'M1' macrophage phenotype (Table 5 (Appendix), Figure 27A).

To extract transcriptional dynamics using Timestamps, I first identified genes that are differentially aged: I calculated the age of each transcript using the maximum likelihood model and then identified all genes for which the distribution of maximum likelihood estimates differs significantly between the 0h and 6h conditions.

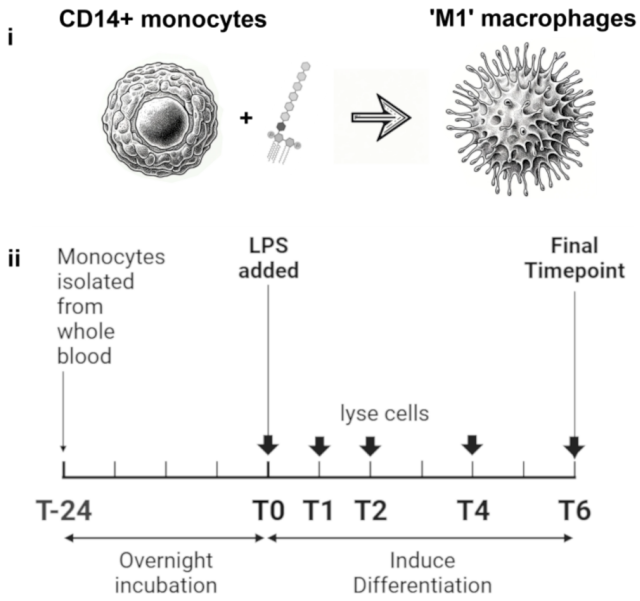
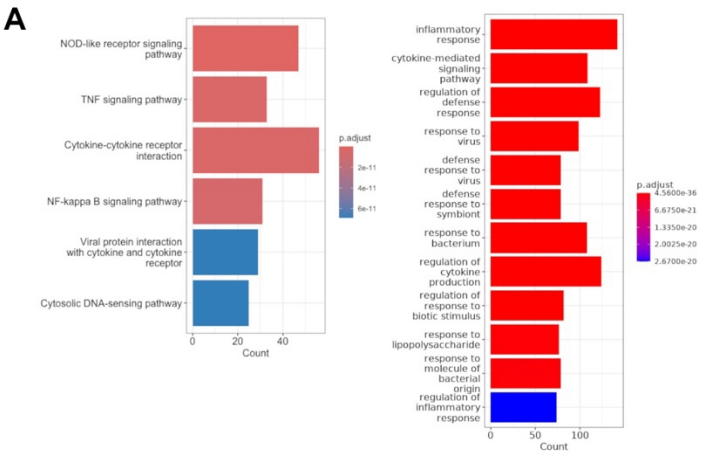
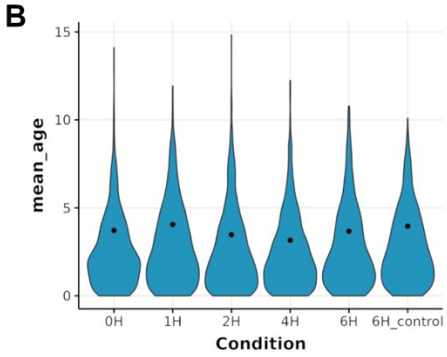


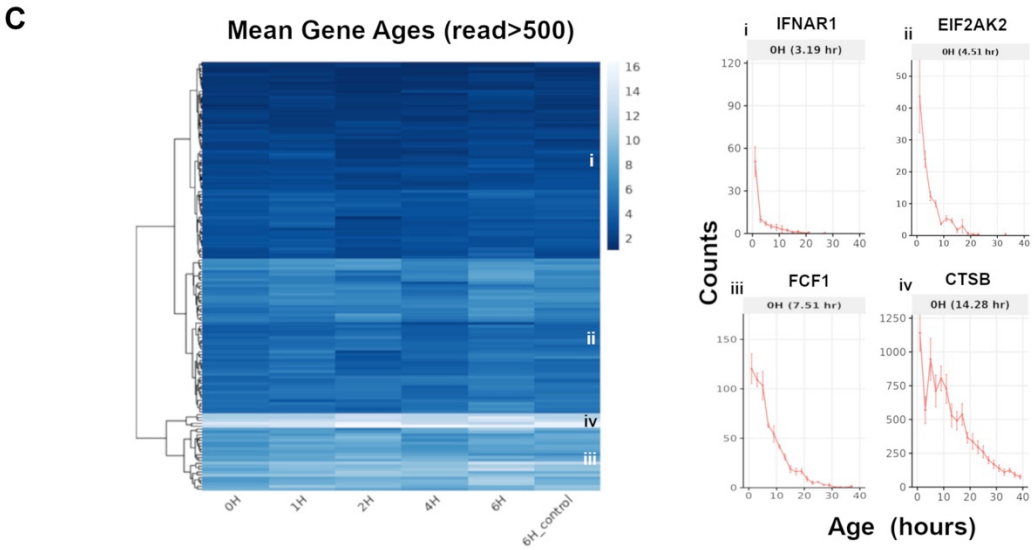
Figure 26 Human primary monocytes were extracted from whole blood and stimulated with E. coli lipopolysaccharide at 0H, before wells were lysed in triplicate at the timepoints shown (cell sketches produced by ChatGPT (OpenAI)).

12.2. Timestamps captures the age of swathes of the human transcriptome

After filtering on sequencing coverage and number of suitable editing sites (see Methods for Chapter 12), I measured the ages of 763 genes at each timepoint. The mean age of these genes in the monocytes was 3.7 hours at T0, with little global change between timepoints (Figure 27B). As expected, most human transcripts are less than 5 hours old (Herzog et al. 2017). There are substantial inter-gene differences in mean age, likely a result of differential transcript stability. For instance, while some genes exist almost exclusively as very young transcripts (mean age <1 hour old), others are veritable Methuselahs, with mean ages > 10 hours (Figure 27C). Gene age distributions generated by Timestamps are sensitive to bursts of transcription, however the gene decay rates appear to be a relatively stable property of the gene (Figure 27C and D).


A
NOD-like receptor signaling pathway
TNF signaling pathway
Cytokine-cytokine receptor interaction
NF-kappa B signaling pathway
Viral protein interaction with cytokine and cytokine receptor
Cytosolic DNA-sensing pathway
Count
p.adjust
inflammatory response
cytokine-mediated signaling pathway
regulation of defense response
response to virus
defense response to virus
defense response to symbiont
response to bacterium
regulation of cytokine production
regulation of response to biotic stimulus
response to lipopolysaccharide
response to molecule of bacterial origin
regulation of inflammatory response
Count
p.adjust


B
mean_age
Condition
0H
1H
2H
4H
6H
6H_control


C
Mean Gene Ages (read>500)
i
ii
iii
iv
0H
1H
2H
4H
6H
6H_control
IFNAR1
EIF2AK2
FCF1
CTSB
Counts
Age (hours)

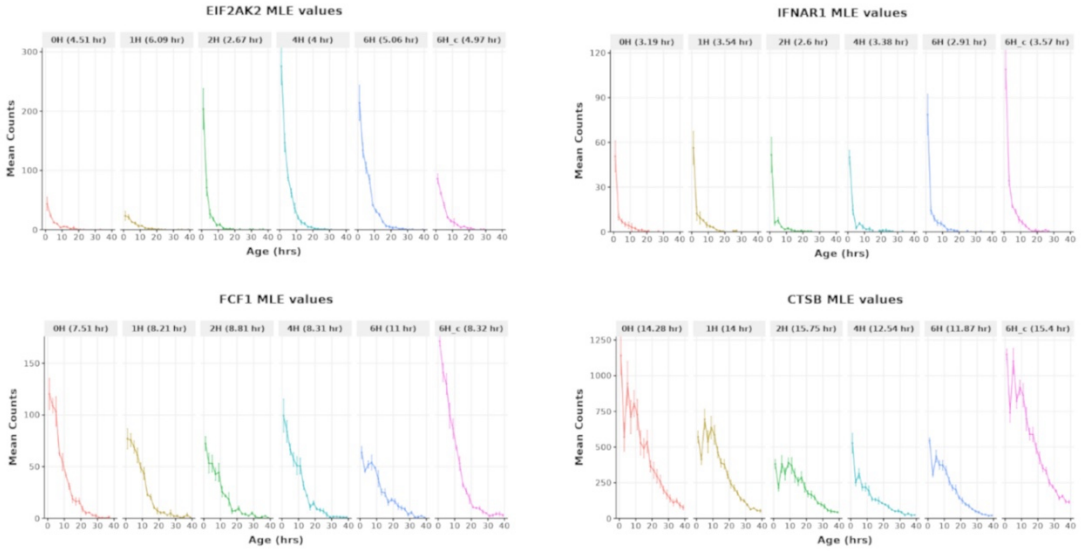

D
EIF2AK2 MLE values
IFNAR1 MLE values
FCF1 MLE values
CTSB MLE values
Mean Counts
Age (hrs)

Figure 27 Timestamps measures the age of hundreds of genes across the human monocyte transcriptome

(A) Gene enrichment analysis (pathways (left) and gene ontology terms (right)) for comparing monocyte gene expression 6 hours after LPS stimulation to baseline expression.
(B) Mean gene age of the 763 genes which could be measured by Timestamps (n = 3).
(C) Mean gene ages of a subset of the 763 genes where at least 500 reads were mapped. (Left) Heatmap – shows the distribution of transcript ages in the monocyte transcriptome, across the samples timepoints, clustered by mean age. The top four clusters identified by hierarchical clustering are identified I – IV. (Right) mean age histograms of 4 representative genes, each from respective clusters I – IV identified on the heatmap. The age histograms are taken from baseline (0H timepoint). (n = 3; bin widths = 2.5 hours; SEM shown).
(D) The age histograms of 4 genes are shown. For each gene, histograms for the 0 – 6H timepoints and the 6H control are shown. (n = 3; bin widths = 2.5 hours; SEM shown).

12.3. Differential transcript age identifies gene regulation not captured in differential abundance data.

As shown in Chapter 11, mean gene age is sensitive to transcriptional activation/shutdown. I wondered if changes in mean transcript age between timepoints would reveal transcriptional changes which are not detected in differential expression analysis. I performed differential expression and differential age analysis (Methods for Chapter 12) on the 0H and 6H timepoint data. Out of 255 differentially expressed genes, I found 10 genes which are differentially aged but not differentially expressed between the 6 hour and baseline (Table 3, Figure 28A), none of which are known to have specific roles in monocyte differentiation (Reed et al. 2022). In general, these genes appeared to undergo transient transcriptional changes such as activations or deactivations that leave signatures in the age of the overall population without changing the bulk levels of expression. For example, RPS19 and RPL37A, two ribosomal proteins, undergo a significant increase in abundance at the 2h timepoint but return to their initial expression levels by the 6h timepoint (Figure 28B). This transient increase in expression is not detectable by differential expression at the 6h timepoint. By contrast, the Pex13 gene undergoes a sharp and highly significant decrease in expression at the 1h timepoint before recovering by the 6h timepoint. This change in

transcription leaves a change in mean transcript age, detectable by Timestamps, while abundance has returned to baseline.

To my knowledge, none of these ten genes are known to have specific roles in monocyte to macrophage differentiation (Reed et al. 2022). In fact, RPL37A, a gene encoding a ribosomal protein, has previously been identified as a housekeeping gene for qPCR studies of monocyte to macrophage differentiation due to its perceived stable expression (Maeß, Sendelbach, and Lorkowski 2010). Using Timestamps, I can potentially uncover new roles for genes which are critical in development and disease. Due to their transient activation, these genes may be overlooked in differential expression analysis and other techniques which take only instantaneous measurements of the transcriptome.

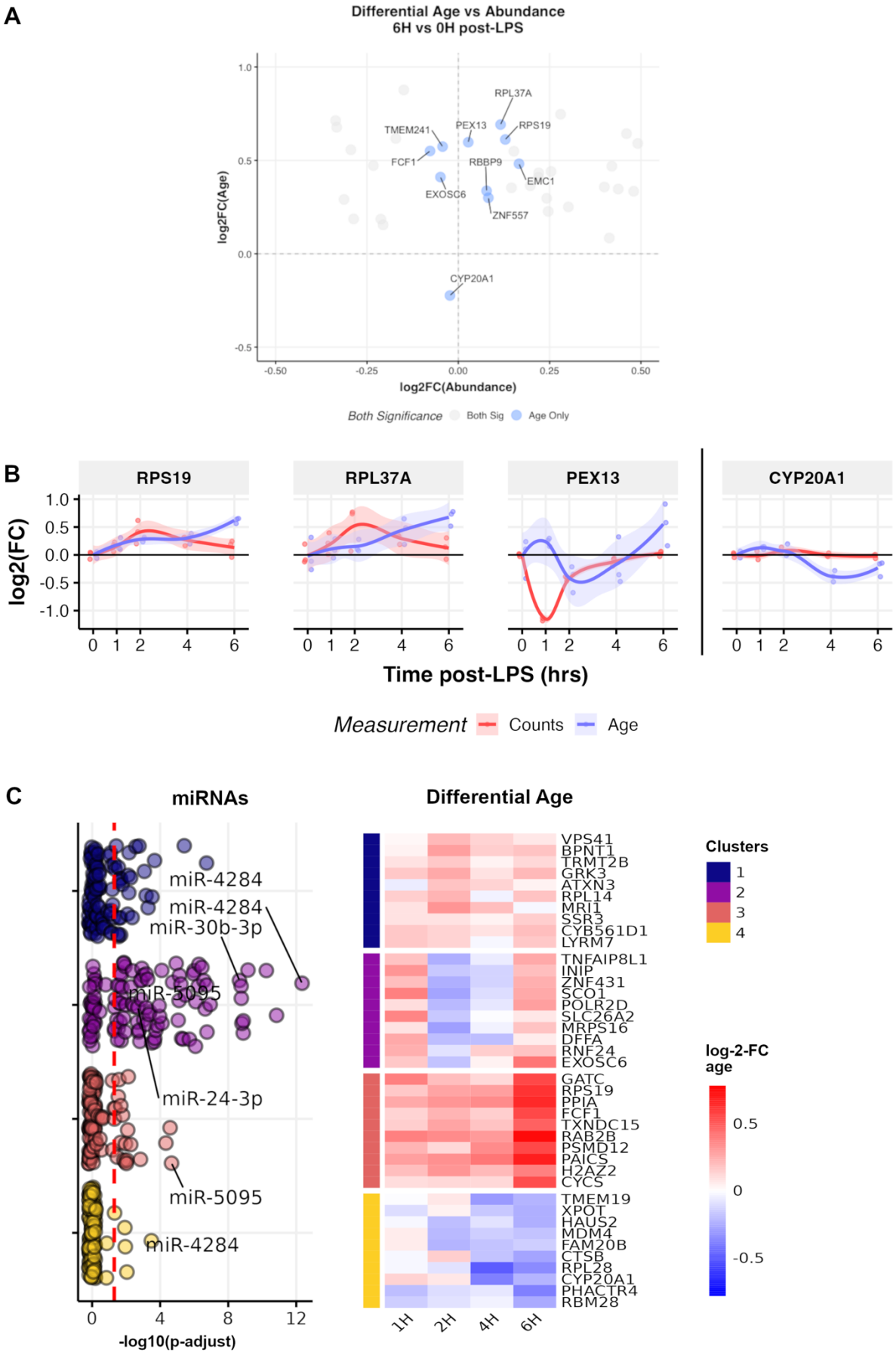

A
Differential Age vs Abundance
6H vs 0H post-LPS
log2FC(Age)
log2FC(Abundance)
RPL37A
RPS19
PEX13
TMEM241
FCF1
RBBP9
EMC1
EXOSC6
ZNF557
CYP20A1
Both Significance
Both Sig
Age Only
B
RPS19
RPL37A
PEX13
CYP20A1
log2(FC)
Time post-LPS (hrs)
Measurement
Counts
Age
C
miRNAs
-log10(p-adjust)
miR-4284
miR-4284
miR-30b-3p
miR-5095
miR-24-3p
miR-5095
miR-4284
Differential Age
VPS41
BPNT1
TRMT2B
GRK3
ATXN3
RPL14
MRI1
SSR3
CYB561D1
LYRM7
TNFAIP8L1
INIP
ZNF431
SCO1
POLR2D
SLC26A2
MRPS16
DFFA
RNF24
EXOSC6
GATC
RPS19
PPIA
FCF1
TXNDC15
RAB2B
PSMD12
PAICS
H2AZ2
CYCS
TMEM19
XPOT
HAUS2
MDM4
FAM20B
CTSB
RPL28
CYP20A1
PHACTR4
RBM28
1H
2H
4H
6H
Clusters
1
2
3
4
log-2-FC age
0.5
0
-0.5

Figure 28 Timestamps identifies transcriptional regulation not found by differential abundance analysis

(A) Comparison of differential abundance versus differential expression analysis for the 6H versus 0H timepoint. Ten genes which are identified as differentially ages but not differential expressed are highlighted in blue. Significance is determined using DESeq2 (12.7.5.1) for abundance ($p<0.05$) and the Wilcoxon signed-rank test for mean age ().

(B) Plots show the relationship between counts (red) and age (blue) across the sampled timepoints for four genes which were differentially aged but not differentially expressed. RPS19, RPL37A and PEX13 undergo transient changes in abundance between the 0H and 6H timepoints.

(C) Genes were clustered by differential age relative to the 0H timepoint. To display the data, four clusters of interest are shown in a heatmap of differential age. Differential age is shown as the log-2-fold change. The centroids of these 4 clusters were calculated and the 10 genes with the shortest Euclidean distance to the centroid are shown. G:profiler (Kolberg et al. 2023) was used to identify miRNAs and transcription factors (TFs) enriched with each cluster (Methods; Table 6). Jitter plots to the left of the heatmap display the -log10(p-adjust) values of enriched miRNAs and TFs. The red dashed line shows significance ($p<0.05$). In the miRNA plot, the top 3 miRNAs by significance are labelled.

Table 3 Genes identified in differential age analysis.

Gene Name	shrunkLFC (abundance)	p_adjust	log2FC (age)	adj_p_value
EMC1	1.66E-01	1.38E-01	4.82E-01	2.36E-04
CYP20A1	-2.28E-02	7.02E-01	-2.23E-01	1.41E-05
PEX13	2.73E-02	7.00E-01	5.97E-01	2.34E-03
RPL37A	1.16E-01	4.52E-01	6.92E-01	5.83E-04
FCF1	-7.74E-02	1.22E-01	5.50E-01	1.47E-27
EXOSC6	-4.88E-02	7.98E-01	4.11E-01	5.53E-03
TMEM241	-4.30E-02	7.36E-01	5.74E-01	1.82E-02
RPS19	1.29E-01	1.91E-01	6.12E-01	2.09E-11
ZNF557	8.23E-02	2.64E-01	3.00E-01	3.23E-02
RBBP9	7.74E-02	2.68E-01	3.37E-01	1.94E-02

12.4. Gene age distributions encode the cell's recent transcriptional history.

I used the sequencing data from intermediate timepoints to identify the transcriptional changes which caused the changes in mean age between the 0H and 6H timepoints. I found multiple genes which are detected only in the differential age analysis undergo transient yet significant ($p<0.05$) changes in abundance at intermediate timepoints and then return to baseline at 6H. These changes cannot be detected in the differential expression analysis (6H versus 0H) but leave legible changes in the gene's age-distributions. For example, RPS19 and RPL37A are significantly differentially expressed at the 2H but not 6H timepoint. RPS19 and RPL37A, both ribosomal proteins, undergo a transient increase in abundance, peaking at 2H before returning to the 0H levels at 6H (Figure 28B). In revealing transient transcriptional changes which are missed in differential expression analysis, I find confirmation that Timestamps uncovers a cell's recent transcriptional history.

12.5. Gene age distributions uncover post-transcriptional regulation.

Intuitively, a change in mean age would be associated with a change in abundance (although the relationship between abundance and age is complicated, as I discovered in Chapter 12). CYP20A1 is of particular interest because its decrease in mean age is not associated with any change in abundance (Figure 28B). Therefore, I reasoned that the shift towards younger transcripts must be caused by an increase in degradation of the transcript. In support of this, CYP20A1 transcripts are predicted to harbor at least 10 miRNA recognition elements, and miRNA overexpression experiments suggest that CYP20A1 transcript stability is indeed regulated by miRNAs (Bhattacharya et al. 2021; Singhal, Dhamija, and Mukerji 2023). Additionally, monocytes treated with LPS for 6 hours are known to upregulate 7 miRNAs which are predicted to bind to the 3' UTR of CYP20A1 (Lu et al. 2019).

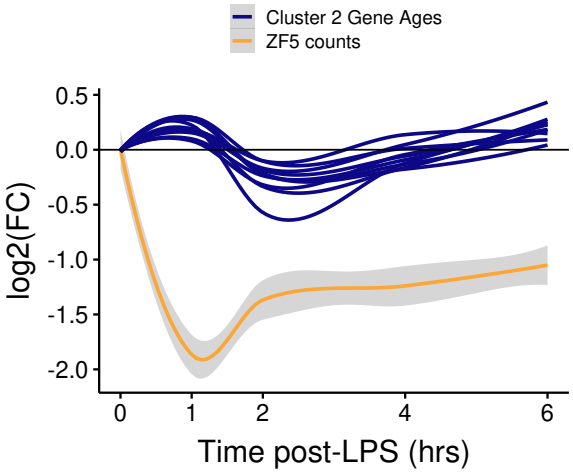


Figure 29 log2 fold change in expression of ZF-5 and genes in cluster 2 (standard error shown in grey for ZF-5 only for clarity).

12.6. Differential gene age identifies new transcriptional modules.

Since mean age seems to be sensitive to both changes in transcription and degradation, I reasoned that clustering genes by differential age between timepoints would reveal new 'transcriptional modules' – miRNA-TF combinations which control mRNA turnover. Now utilising data from all the experiment timepoints, I clustered genes by mean age relative to baseline. I then identified miRNAs and TFs which were enriched in the clusters using G:profiler (Figure 28C; Table 6; Sections 12.7.5.3 and 12.7.5.4).

Identification of TFs enriched in co-differentially aging genes identified ZF-5 (*ZBTB14*) which is known to regulate the differentiation of monocytes in zebrafish, but whose role in human monocyte differentiation is not known. In zebrafish, ZF-5 represses the transcription of PU.1, a master regulator of monocyte differentiation and putative pioneer TF (Lawrence and Natoli 2011), thereby limiting monocyte proliferation. In support of a similar function in humans, a loss-of-function mutation in the *ZBTB14* gene ($ZBTB14^{S8F}$) has been identified in a patient with Acute Myeloid Leukaemia (Tyner et al. 2018). (Lawrence and Natoli 2011). In my analysis, all 88 genes in cluster 2 are in proximity to a ZF-5 regulatory motif, suggesting that Cluster 2 represents a core transcriptional motif for monocyte differentiation. Additionally, I found that changes in expression of ZF-5 closely tracks that of cluster 2 genes at the sampled timepoints (Figure 29).

Timestamps also identified numerous microRNAs that are likely to be involved in the monocyte to macrophage differentiation program. Micro RNAs have a well-established role in monocyte differentiation, modulation of the inflammatory response and in determination of cell fate in general. My analysis identified miR-4284 as a regulator of clusters 1-4, which is known to be upregulated in human monocytes treated with LPS. My analysis also identified the miR-30b/miR-24 duo in cluster 2, which are both known to temper the production of cytokines and phagocytic capacity in human macrophages (Naqvi, Fordham, and Nares 2015; Fordham, Naqvi, and Nares 2015). Finally, I also identify several novel miRNA modulators of monocyte differentiation (Figure 28C). For example, miR-5095 is the most

highly enriched miRNA in cluster 3 and has not previously been associated with human monocyte differentiation. Surprisingly, the miR-5095 mouse ortholog belongs to the super-enhancer of genes which determine CD14+ monocyte cell identity, suggesting it may play a role in human monocyte-to-macrophage differentiation (Hnisz et al. 2014).

12.6.1. Estimation of error in MLE of single molecule age

To calculate an explicit estimate of error in single molecule age, I partitioned the monocyte sequencing data (See only the steady state samples) into a training set (replicates 1 and 2) and a test set (replicate 3) to assess the performance of the maximum likelihood estimation (MLE) of mRNA molecule age. The training set was used to compute likelihood functions and derive MLE values for each molecule, and an average age per gene was calculated to serve as an internal benchmark. Subsequently, the model was applied to the test set, and the estimated ages were compared against the corresponding training-set averages to quantify prediction errors. Error metrics, specifically the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), were computed, and the results were visualized in a scatter plot of estimated versus benchmark ages, thereby enabling an assessment of model consistency and reliability in estimating mRNA molecule age. The strong correlation between the training and test set gives confidence that the estimates are reliable.

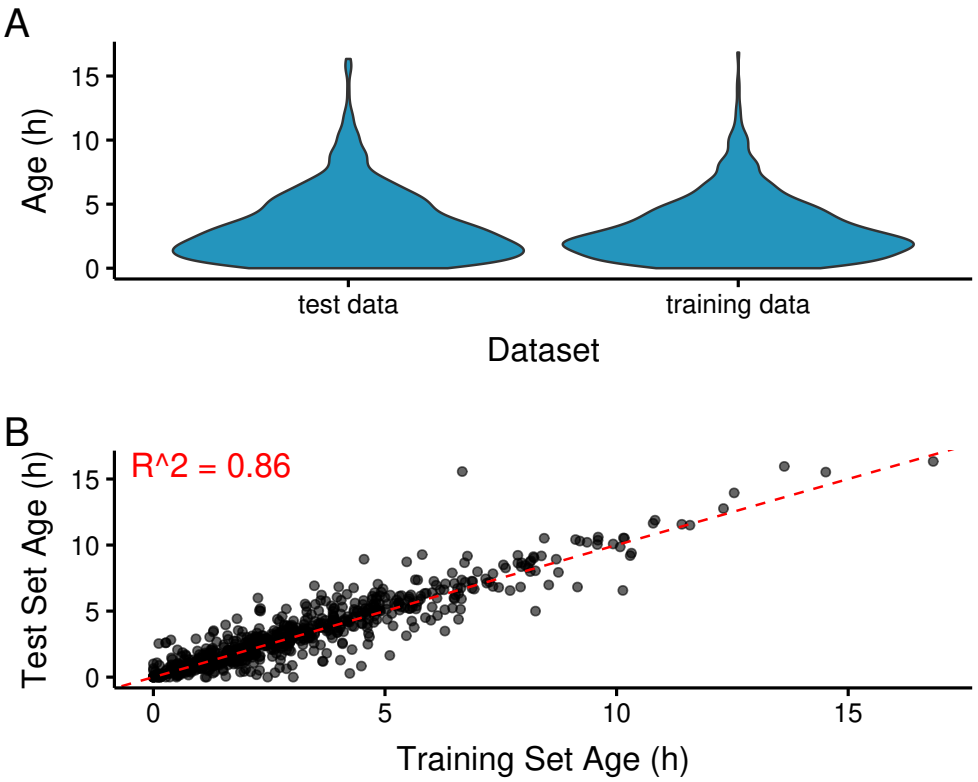


Figure 30 Estimation of MLE error using a train-test split.

(A) Violin plots show the single-molecule age distributions from the MLE in the training and test datasets.

(B) Scatter plot of per-gene (n=727 genes) mean ages in the training versus test datasets. R^2= 0.86

12.7. Methods for Chapter 12

12.7.1. Peripheral Blood Mononuclear cell (PBMC) isolation

Whole blood was sourced from a healthy male via the UK National Health Service blood bank. Blood volume was standardized by supplementing with Hank's Balanced Salt Solution (HBSS) containing 0.3mM EDTA to a final volume of 25ml. Next, 20ml of Ficoll-Paque was dispensed into separate 50ml tubes. The blood was delicately layered atop the Ficoll using a 5ml pipette, ensuring the preservation of the discrete interface. The sample was centrifuged at 700g, without brake, for 25 minutes at room temperature. Post-centrifugation, the PBMCs, identifiable as a white interphase layer, were harvested using a Pasteur pipette and transferred into a clean 50ml tube. The cell suspension was diluted to 50ml with HBSS 0.3mM EDTA, followed by a centrifugation step at 500g for 10 minutes at room temperature. The supernatant was decanted, and the cell pellet was subsequently resuspended in 40ml HBSS 0.3mM EDTA. This wash step was repeated once. After the final centrifugation, the PBMCs were resuspended in 1ml of MACS buffer in preparation for CD14+ isolation.

12.7.2. CD14+ cell isolation

Cells were processed using the MagniSort system (Invitrogen) according to the manufacturer's protocol. Initially, 200 µl of MagniSort Enrichment Antibody was added to the cell suspension and thoroughly vortexed. The mixture was then incubated at room temperature (RT) for 10 minutes. Following this, 3 ml of MACS buffer was added, and the sample was centrifuged at 300g for 3 minutes at RT. The resulting pellet was resuspended in 1 ml of MACS buffer.

Next, 300 µl of MagniSort Positive Selection Beads was added to the suspension, and the solution was vortexed to ensure even distribution. After a 10-minute incubation at RT, the volume was adjusted to 2.5 ml with MACS buffer, followed by gentle mixing through pipetting. The tube was then placed on a magnetic separator and incubated at RT for 5 minutes to allow for magnetic separation. The supernatant, containing CD14- cells, was

carefully discarded. The magnetic separation process was repeated twice more, each time resuspending the beads in fresh MACS buffer.

The retained CD14+ cells were then resuspended in 30 ml of Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% Fetal Bovine Serum (FBS), 1% Penicillin-Streptomycin-Glutamine, and 100 ng/ml M-CSF to support cell survival in the 24 hours before the experiment.

12.7.3. LPS induction of human monocytes

Human CD14+ monocytes were seeded in triplicate into 6-well plates at a density of 2 million cells per well immediately after magnetic isolation. The cells were allowed to adhere and rest for 24 hours prior to stimulation. At time t=0 hours, lipopolysaccharide (LPS) derived from *Escherichia coli* and purified by ion-exchange chromatography (Sigma, L3024) was added to each experimental well to achieve a final concentration of 100 ng/ml.

12.7.4. Library preparation and sequencing

Samples were reverse transcribed and amplified using the same protocol as described for the HEK293 induction experiment (refer to the protocol outlined above). Sequencing was performed on a PromethION24 platform (ONT) using eighteen R10.4.1 flow cells, with one flow cell dedicated to each sample. This approach yielded an average depth of approximately 70 million reads per flow cell, resulting in a total of over 1.2 billion long reads.

12.7.5. Modelling and Statistics

12.7.5.1. Differential Expression

We initially constructed a DESeqDataSet object (using the R package `DEseq2`) from our isoform count matrix and corresponding sample annotations to define the experimental design (~Condition). Then, by applying the `multi_deseq` function with a set of predefined

contrasts and enabling shrinkage, we obtained more reliable differential expression estimates. This approach helps to reduce noise in log fold change estimates, particularly benefiting the analysis of lowly expressed isoforms.

12.7.5.2. Differential Age

To evaluate differential age between two conditions, I calculated the log2 fold change (log2FC) in mean age and assessed significance using a Mann-Whitney U test on the pooled MLE values. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method.

I plotted the log2FC in both transcript age and counts for each replicate relative to the mean value of the 0-hour replicates. The data were displayed as points with a smoothed LOESS line, and the shaded area indicated 95% confidence intervals.

12.7.5.3. Clustering

For clustering analysis, genes were grouped based on their differential age relative to the 0-hour time point, resulting in 10 distinct clusters, 4 of which are shown for display purposes. Figure 28 highlights four clusters of interest, showing the centroids and the 10 genes with the shortest Euclidean distance to each centroid. Cluster robustness was tested using a non-parametric bootstrap method. After filtering for robustness, I retained 4 out of the initial 10 clusters. In each bootstrap iteration, the dataset was resampled, and clustering was repeated on the resampled data. Each original cluster was compared to the most similar cluster from the bootstrapped data, and similarity was calculated. Clusters that maintained their structure after bootstrapping were considered robust (Methods).

12.7.5.4. Enrichment Analysis

G:Profiler (Kolberg and Raudvere 2024) was used to identify miRNAs and transcription factors enriched within each cluster, with a background set of all genes with more than 5 reads in the monocyte RNA-seq data (full results shown in Table 6 - Appendix). Multiple

testing correction was conducted using the g:SCS (g:Profiler: Set, Counts and Sizes) threshold.

13. Timestamps extends the capabilities of Single Cell RNA Sequencing.

Given that past transcriptional events are encoded in gene age distributions, I wondered if this phenomenon could be used to extract dynamic information from single cell RNA sequencing data. Currently, a majority of single cell RNA sequencing tools, whether conventional or spatial, are used for cell classification, dimensionality reduction or clustering (Zappia & Theis 2021). These categories represent a large and growing proportion of scRNAseq tools, while tools which extract information about cell activity, such as velocity methods represent a diminishing proportion of new scRNAseq tools. This limitation is largely imposed by the relative low sensitivity of gene detection in RNA sequencing methods. Even in single cell in plate sequencing, only a fraction of the transcriptome is captured, usually the most highly expressed genes. This is approximately 6000 genes per cell in SMARTseq2 and 3000 genes per cell for 10X sequencing (Wang et al. 2021). Transcript age distributions could provide orthogonal information from single cell RNA sequencing data about past cell states, thereby enriching the information obtained from single cell RNA sequencing, without an increase in sensitivity.

I applied Timestamps to RNA-seq data acquired using the SMARTseq2 protocol on dividing single HEK-293 cells in well plates. When dividing cells are sequenced in bulk, unless cell division has been paused, the sequencing data will resemble the mean of an ensemble of different cell cycle states. Single cells RNA sequencing allows us to resolve these states (Tirosh et al. 2016).

13.1. Cell cycle phase is encoded by mRNA age.

I used the regular cell cycle of HEK293 cells as a metronome to study the ability of Timestamps to recover the expected dynamics of canonical cell cycle associated genes. I used the gene expression of 96 canonical cell cycle genes to order HEK293 cells based on

whether they were most likely to be in S or G2/M-phase (Figure 31) (Tirosh et al. 2016). I identified the top 15 S- and M-phase cells and explored the differences in the ages of canonical cell cycle genes in those cells (Figure 32). In general, most genes which could be measured by Timestamps were younger in S-phase ((Figure 32A). This is consistent with a surge in global transcription taking place in S-phase (Meryet-Figuiere et al. 2014). (It is also possible that the subset of genes which are measured by Timestamps were specifically more highly expressed in S-phase, although this is unlikely because GO analysis did not show any enrichment in these genes compared to all genes which could be detected by RNA-seq).
Consistent with the role of *HELLS* as a member of the SNF/RAD54 helicase family that maintains genome stability during DNA replication, I noted that HELLS was relatively younger in S- versus G2/M-phase cells (Figure 32A and B). Additionally, I noted that TMPO (a G2/M-phase associated gene) was younger in G2/M-phase cells than S-phase cells, consistent with its expression in G2/M phase ((Figure 32B) (Zhang et al. 2016).

I then investigated whether the cyclic gene expression patterns corresponding to the phases of the cell cycle could be inferred from Timestamps-derived age estimates. After filtering, I used non-linear least squares regression to fit a sine curve to the age histograms (Figure 32C; Table 7). Bootstrapping this approach through 750 iterations (Methods for Chapter 13) revealed a cyclical expression pattern in HELLS ($p=0.044$), with a period of 12.4 hours. *MDM2* was also expressed in a cyclical fashion ($p=0.047$), with a period of ~14.8 hours, with transcripts less than ~7 hours more prevalent in S-phase cells, and those aged ~7-14 hours more prevalent in M-phase cells. This is consistent with known roles of *MDM2*, an oncogene that responds to DNA damage that might be generated during S-phase (Jin et al. 2008; Bouska and Eischen 2009).

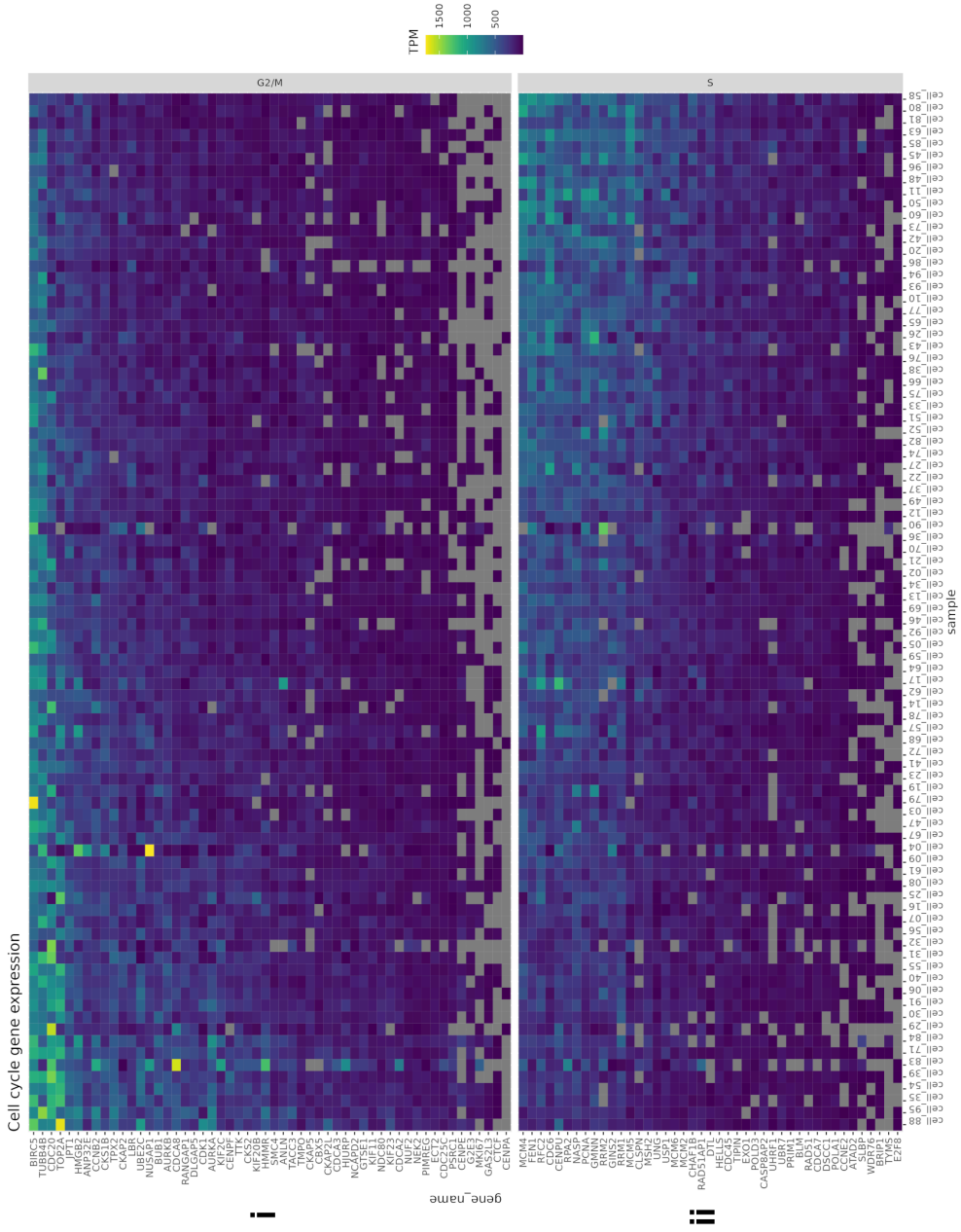
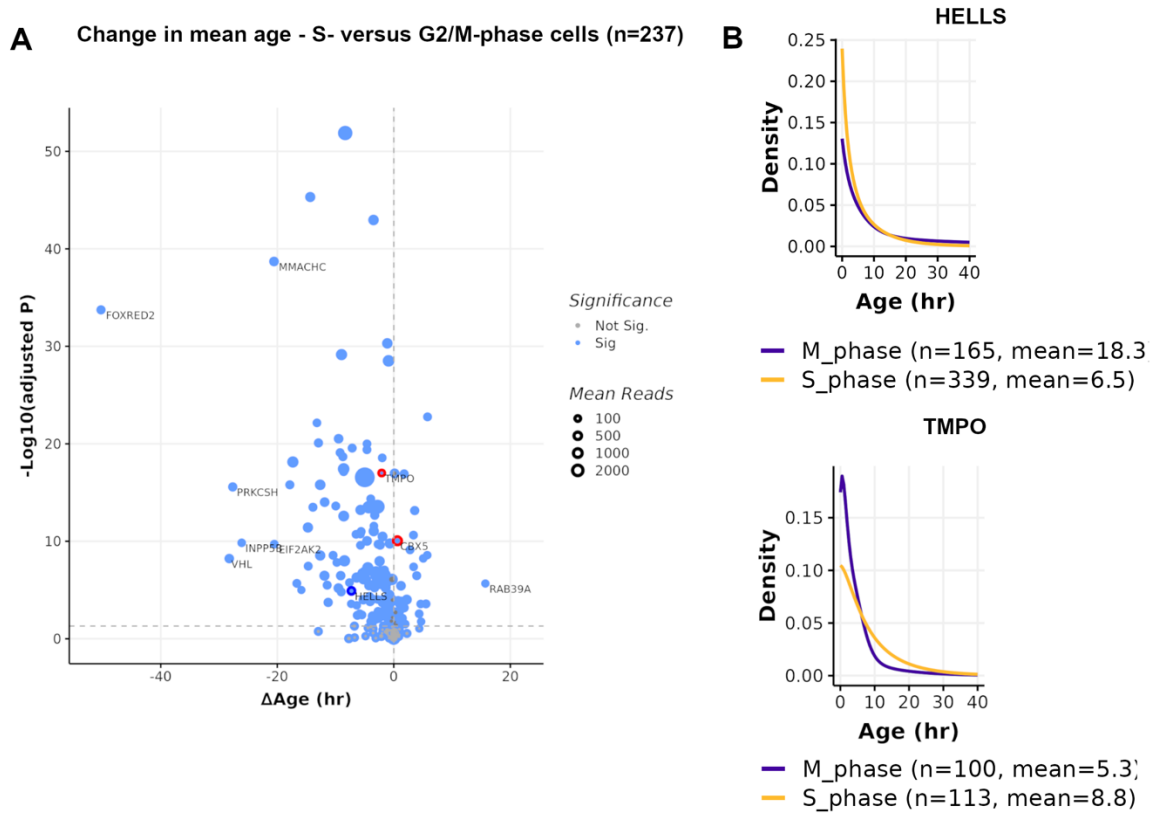
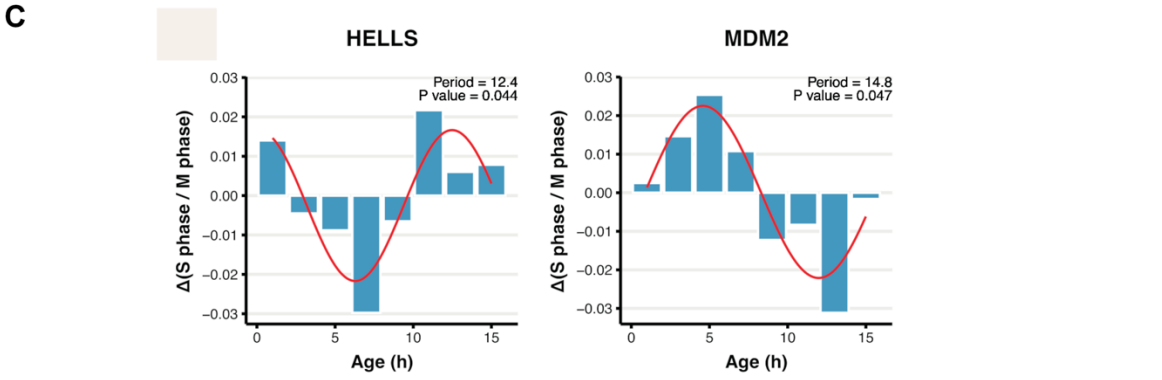


Figure 31 HEK293 cells ranked by increasing likeness to G2/M (i) or S-phase (ii)

HEK293 cells were sorted into wells of a 96-well plate, prepared according SMARTseq2 and then sequenced with Nanopore sequencing. As per Tirosh et al., cells were ranked according to the ratio of expression of G2/M and S-phase genes (Tirosh et al. 2016). Panel (i) shows HEK293 cells ranked by increasing G2/M phenotype from right to left). Panel (ii) shows cells ranked by increasing S-phase phenotype (from left to right).


A
Change in mean age - S- versus G2/M-phase cells (n=237)
-Log10(adjusted P)
ΔAge (hr)
FOXRED2
MMACHC
PRKCSH
INPP5E
EIF2AK2
VHL
TMPO
TBX5
HELLS
RAB39A
Significance
Not Sig.
Sig
Mean Reads
100
500
1000
2000
B
HELLS
Density
Age (hr)
M_phase (n=165, mean=18.3
S_phase (n=339, mean=6.5)
TMPO
Density
Age (hr)
M_phase (n=100, mean=5.3)
S_phase (n=113, mean=8.8)


C
HELLS
Period = 12.4
P value = 0.044
Δ(S phase / M phase)
Age (h)
MDM2
Period = 14.8
P value = 0.047
Δ(S phase / M phase)
Age (h)

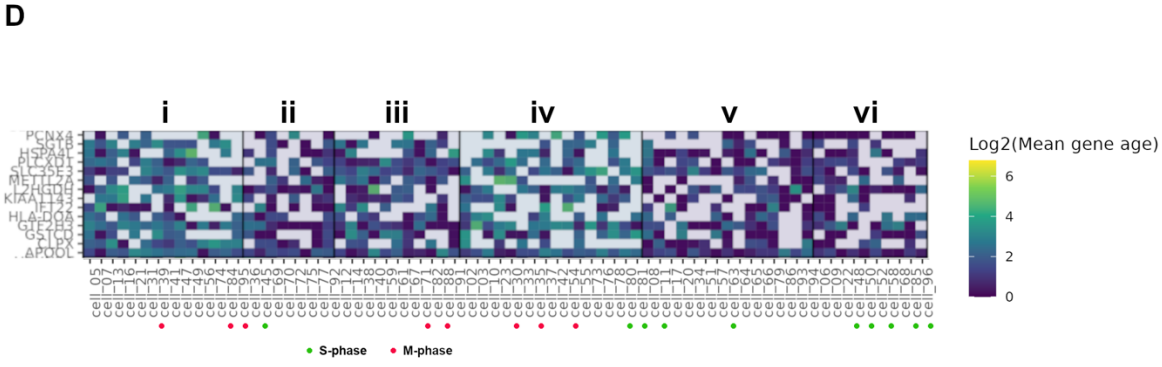

D
i
ii
iii
iv
v
vi
Log2(Mean gene age)
S-phase
M-phase

Figure 32 Timestamps measures the ages of single molecules of mRNA in single HEK293 cells

(A) Scatter plot of significance versus change in age between the top 15 S- and top 15 G2/M-phase cells (filtered to display the 237 genes with >100 reads).

(B) MLEs for HELLS (an S-phase gene) and TMPO (an M-phase gene) in M- and S-associated cells.

(C) Transcript age histograms where each age bin shows the change in density in S- versus G2/M-phase cells. (Data for the top 15 S- versus G2/M-phase cells). A sine curve is fitted in red. The periodicity of the curve in hours and p-value of the bootstrap is shown (See section 13.4.2 for details).

(D) A heatmap produced by biclustering cells and genes based on mean gene age for all genes for which ages could be measured using Timestamps (subset of the heatmap shown). Clusters i-vi are discussed in the text. The top 10 S- and G2/M-phase cells are shown by green and red dots respectively.

13.2. Timestamps interrogates continuous biological processes at the level of single cells.

To further explore transcriptional networks involved in dividing HEK293 cells, and noting the cyclical expression of transcripts, I filtered transcripts to those less than 4 hours old and clustered them using a biclustering approach due to the sparsity of the data (Figure 32D shows a subset of the total heatmap) . Filtering the transcripts less than 4 hours old allowed us to increase the association of transcripts with the respective phase of the cell cycle in which expression is initiated by removing latent (older) transcripts which are left over from a previous cell cycle phase. Using this approach, I was able to separate S-phase from G2/M-phase cells. Figure 32D shows that S-phase cells clustered predominantly into clusters v and vi. While I clustered all genes for which an age could be determined by Timestamps (554 genes in total), I have displayed only 10 genes in Figure 32D. This suggests that in single cell RNAseq, transcript ages measured using Timestamps can be used to infer real biological processes.

Taken together, these results show that Timestamps can reveal dynamical information about RNA expression not present in traditional RNA sequencing analysis, which provides only a snapshot of a cells behaviour.

13.3. The limitations of Timestamps in analysing single cell RNA sequencing data

Some limitations of this analysis of single cell data must be noted here. Timestamps is inherently limited to analysing genes which have enough editing sites (each with sufficient editing rates) in their 3' UTR. Furthermore, only genes which are present in the calibration experiment can be studied with Timestamps. If the calibration was performed in a different cell line to the discovery experiments, which was the cause for my analysis of human monocyte differentiation in the previous chapter, then monocyte-specific genes cannot be analysed. This could be overcome by performing a calibration experiment on the

experimental cell line, especially if this was known in advance. In this case, a monocyte calibration with 4 timepoints and 3 replicates would have been sufficient for fitting a CDF to the editing data, however, I opted to use the HEK293 calibration for budgetary reasons.

Even if the cell line used in the calibration and discovery experiments are the same, genes may be upregulated in the context of the discovery experiment which are not present in the calibration. For example, this would be the case for genes which are expressed when HEK293 cells are subjected to heat shock but in not in HEK293 cells at baseline.

With HEK293 single cell data, the analysable genes are subject to both the constraints on number and editing rates of sites in the 3' UTR and the low sensitivity of single cell RNA sequencing. The number of detected editing sites is reduced by the relatively low sequencing depth compared to bulk sequencing (See Section 8.2.2). When interrogating genes associated with a specific cell process, such as cell division, the number of genes at the union of sufficient editing and sufficient depth is small. For example, in the above analysis I found that only one S-phase (HELLS) and one G2/M-phase associated gene (TMPO) respectively qualified for analysis with Timestamps. This made discovery analysis beyond verifying the accuracy of the method difficult and prompted further development of hyperactive RNA editors which is continued in chapter 13.4.

13.4. Methods for Chapter 13

13.4.1. Single cell RNA sequencing in plates

HEK293FT cells (Thermo Fisher) were cultured until they reached approximately 70% confluency, then detached via trypsinization, counted, and resuspended in FACS buffer composed of PBS with 2% molecular grade BSA and 1 µg/ml actinomycin D, at a final concentration of 5×10^6 cells per ml. DAPI was added shortly before sorting to serve as a viability dye. Prior to initiating the sorting process, the FACS machine was calibrated using a horseradish peroxidase assay to ensure accurate cell deposition into each well. Single cells were sorted into individual wells of a low-bind 96-well plate (Eppendorf) using a MoFlo XDP cell sorter (Beckman Coulter) fitted with a 100 µm nozzle, utilizing 405 nm lasers and 447/60 filters. The gating strategy involved selecting cells based on forward and side scatter characteristics (FSC/SSC) and employing a doublet exclusion gate, with DAPI-negative cells designated for sorting.

For library preparation, the NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module (NEB, E6421L) was employed to construct SMART-seq libraries. Cells were sorted into wells containing 5 µl of freshly prepared lysis buffer, with reverse transcription and library preparation initiated immediately thereafter. Following reverse transcription, PCR amplification was performed for 22 cycles as per the manufacturer's instructions, followed by an additional 5 cycles to attach nanopore barcodes and adapters. The samples were subsequently pooled and prepared for sequencing using the Ligation Sequencing Kit V14 (ONT, SQK-LSK-114) in accordance with the manufacturer's protocol.

The sequencing libraries were processed on a PromethION24 sequencer (ONT) across six R10.4.1 flow cells, yielding an average output of 44 million reads per flow cell.

13.4.2. Modelling and Statistics

13.4.2.1. HEK293 Single Cell Analysis

Cells with fewer than 1,000 expressed genes or fewer than 20 genes with calibration fit sites were excluded. Cells were ordered based on the S- to M-phase gene expression ratio (Tirosh et al. 2016). The top 15 S-phase and M-phase cells were selected for further analysis. I focused on 554 genes with at least 5 fit sites in the 3' UTR, resulting in a total of 189,707 fitted transcripts. Gene and transcript age likelihoods were calculated as previously described.

To investigate whether gene transcripts might exhibit cyclical expression during the cell cycle, I first calculated the relative density of transcripts for each gene across the top S- and M-phase cells, within 2-hour time bins ranging from 0 to 16 hours. The difference between S-phase and M-phase densities was then computed for each time bin, resulting in a delta-S/M-density-per-bin. My analysis focused on 34 genes with data present in at least 90% of bins and more than 500 total reads across all included cells. Using the non-linear least squares function from the `stats` package, I fitted a sine curve to the delta-densities using the equation:

$$\text{Delta} = A \times \sin(B \times (\text{Age} - C)) + D$$

The starting conditions were: A = 0.05, B = 0.6, C = 0 or 6 if convergence not attained, and D = 0. An upper limit of B=1.57 was set, corresponding to a minimum cell cycle period of 4 hours. For the 25 genes where the model converged within these parameters, the process was repeated with 750 bootstrap iterations, randomly shuffling the S- or M-phase labels in each iteration. Metrics derived included the residual sum of squares and a p-value for the B parameter. The p-value for the residual sum of squares was calculated by counting the number of bootstrap results with lower values than the true residual sum of squares, divided by the total number of bootstrap iterations. The period was then calculated using the equation:

Period = $2\pi/B$.

13.4.2.2. Single Cell Gene Clustering Analysis

For gene clustering, a biclustering approach (J. Li et al. 2020; Reisner, Pham, and Li 2021) was used to group genes by minimizing the total sum of squared errors (SSE) within clusters, iterating 1,000 times and using the Rand index to compare iterations and minimize SSE. Clustering was performed on the mean age per gene using MLEs from all transcripts in the top 15 S- and M-phase cells, followed by filtering to include only transcripts less than 4 hours old.

To assess the impact of filtering transcripts to those less than 4 hours old, I examined the clustering positions of three canonical cell cycle genes, HELLS, MDM2, and TMPO. Functional enrichment of the two clusters containing these genes was analysed using g:Profiler against a background of all genes expressed in the dataset, with multiple testing correction performed using the tailored g:SCS method to correct for multiple testing. The significant results included gene ontology terms and the most significant miRNA and transcription factor associations, which are visualized in Fig. S-7.

14. Improving Timestamps using hyperactive, promiscuous RNA editors – Generation 2.

From my single cell work on HEK293 cells and mouse neurones, it became clear that for Timestamps to be useful in the single cell domain, I would need to develop a promiscuous, hyperactive editor of RNA. The first generation of RNA editors I engineered where based on fusion of the ADAR1(E488Q) catalytic domain, E. coli TadA (ABEmax) or cytosine deaminases.

After reading about the use of a TadA mutant in a novel methylation sequencing method, I became interested in exploring other TadA variants. Evolved TadA-assisted N6-methyladenosine sequencing (eTam-seq), is a method used to detect and quantify m6A modifications in RNA (Xiao et al. 2023). Central to this method is the use of a mutant TadA enzyme, TadA8.20, which facilitates adenosine deamination. TadA8.20 deaminates adenosines (A) but does not act on N6-methyladenosine (m6A), a common RNA modification. This selective activity allows for the detection of m6A sites, as unmethylated adenosines are converted to inosines, which are then read as guanosines during sequencing. In contrast, m6A-modified adenosines remain unedited and are identified as persistent adenosines in the sequencing data. This approach enables transcriptome-wide or site-specific m6A profiling. Although TadA8.20 was used in high concentrations to *ex vivo* deaminate the RNA *("10 ng of RNA was incubated with 10 μM TadA8.20 in 20 μl of deamination buffer at 37 °C for 3 h"),* its ability to act at adenosines throughout the transcriptome was intriguing.

The authors had used the monomeric version of this mutant (TadA8.20m). Tad A natively operates as a dimer (Figure 33). Although both monomers interact with the RNA substrate, one monomer acts as a docking station while the other monomer performs deamination (Gaudelli et al. 2017). A monomeric TadA could potentially show low sequence preference. Furthermore, it is possible to replace one monomer with a peptide which binds RNA

indiscriminately, such as the phage lambda N-protein (which I discovered in section 10.2.5 is highly promiscuous in its binding of RNA).

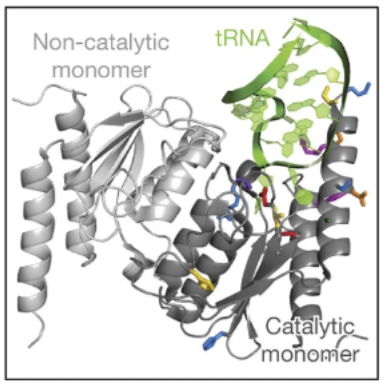


Figure 33 Mechanism of native E. coli deamination of tRNA (Gaudelli et al. 2017)

Work by Gaudelli et al. described another TadA variant, TadA7.10d (a dimer), which showed transcriptome-wide off-target editing (Gaudelli et al. 2020).

14.1. E. coli Dps protein

E. coli Dps (DNA-binding protein from starved cells) is a protein that helps bacteria survive under stress conditions, particularly during the stationary phase. It forms a dodecameric spherical structure that protects DNA from oxidative damage, pH changes, and environmental stress. The N-terminal region of the Dps protein (Dps-N) contains lysine residues that allow binding to the DNA phosphate backbone and condensation of DNA. However, Dps cannot bind DNA during the exponential growth phase of *E. coli*. Instead, it preferentially binds to RNA. The RNA interaction is crucial as it stabilizes the protein and prevents its binding to DNA during non-stress conditions. This suggests that Dps could behave as a non-specific RNA/DNA binding protein (Park et al. 2020).

Work by Kwon and Giessen showed that Dps-N, the N-terminal region of the Dps protein, binds to both RNA and DNA through electrostatic interactions, showing broad affinity for nucleic acids. When engineered and fused to nanostructures such as encapsulins, Dps-N conferred RNA and DNA-binding abilities enabling the construction of protein nanocage RNA delivery vectors (Kwon and Giessen 2022).

This evidence suggested to me that Dps-N could enable broad editing of the transcriptome if fused to a hyperactive editor such as TadA7.10d or TadA8.20m.

14.2. Construction of TadA fusion proteins

Unlike ADAR enzymes, E. coli TadA is potentially not subject to cell feedback mechanisms which act to limit RNA editing. To this end I cloned three constructs: TadA 7.10d, TadA8.20m, and the fusion protein TadA8.20m-Nlambda. Conveniently, a Karl Brune, a post-doctoral researcher in my lab had cloned the fusion protein TadA8.20m-DspN. He had however cloned the construct in a different configuration to my TadA8.20m-Nlambda construct, with Dps-N on the N-terminal side of TadA, and with a shorter linker (Figure 34). Previous work has determined that a GGS linker length of 32 amino acids was optimal for the performance of ABE2.6 (a TadA-Cas9 fusion protein) (Gaudelli et al. 2017). The same paper and others had shown that TadA-Cas9 (N→C) was more effective than Cas9-TadA (Gaudelli et al. 2017; 2020). This had led me to believe that the RNA binding protein (Nlambda or Dps-N) should be placed in the C-terminal side of TadA (thus taking the place Cas9 would occupy). However, the same paper also demonstrated that in TadA-TadA-Cas9 fusions, the internal TadA was responsible for editing, whereas inactivating mutations to the N-terminal TadA had no effect on editing. Given the structural context was very different (Cas9 is approximately 1.4kb, Dps-N is 12 amino acids), it was not if the configuration would matter, and the convenience of the available plasmid, I proceeded with my colleague's Dps-N plasmid without modification (Figure 34).

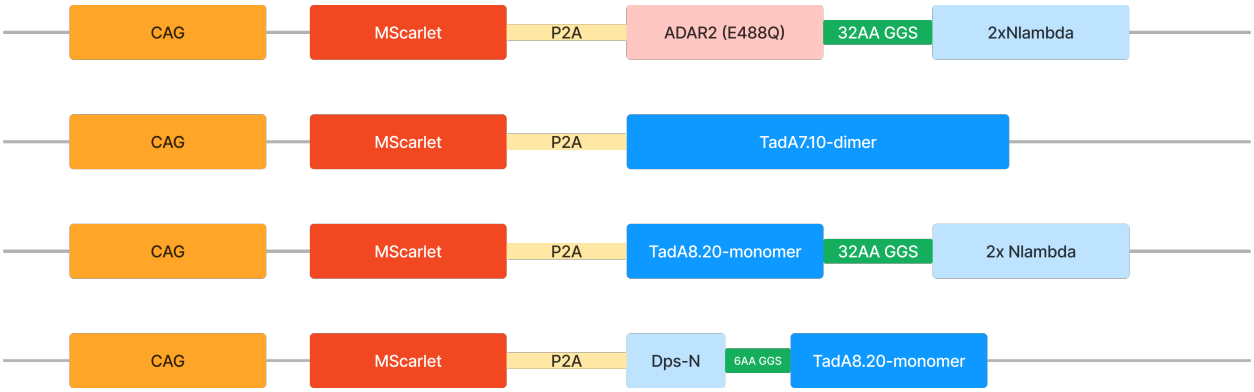


Figure 34 Schematics of the second generation of editors (TadA-based editors) and ADAR2(E488Q)-Nlambda.

14.3. Generation 2 editing construct – transcriptome-wide editing

With the privilege of lessons from the failed editor ADAR2(E488Q)-Nlambda, I made sure to fit an exponential CDF to the new editing sites to calculate the editing rates (Figure 35A). While ADAR2(E488Q)-Nlambda added many additional editing sites, the novel sites were rarely edited and therefore not useful in Timestamps (which form our experience required an editing rate of at least 0.01 edits per hour, but more ideally 0.04 edits per hour). I confirmed this previous finding; ADAR2(E488Q)-Nlambda added few sites with editing rates >0.01. On the other hand, all the TadA constructs showed far higher levels of editing, particularly the TadA8.20-based constructs. Surprisingly, TadA8.20 monomer outperformed TadA7.10 which is dimeric. TadA8.20 was further improved by fusion to an RNA binding partner (Nlambda or Dps-N). The highest level of editing was by TadA8.20-DpsN. It introduced hundreds of thousands of editing sites with an editing rate >0.01 and several thousand with an editing rate >0.04 (Figure 35A).

Comparing the distribution of TadA8.20-DpsN editing across the transcriptome to that of the empty vector control and ADAR2(E488Q)-Nlambda showed it edited across many more genes (Figure 35B).

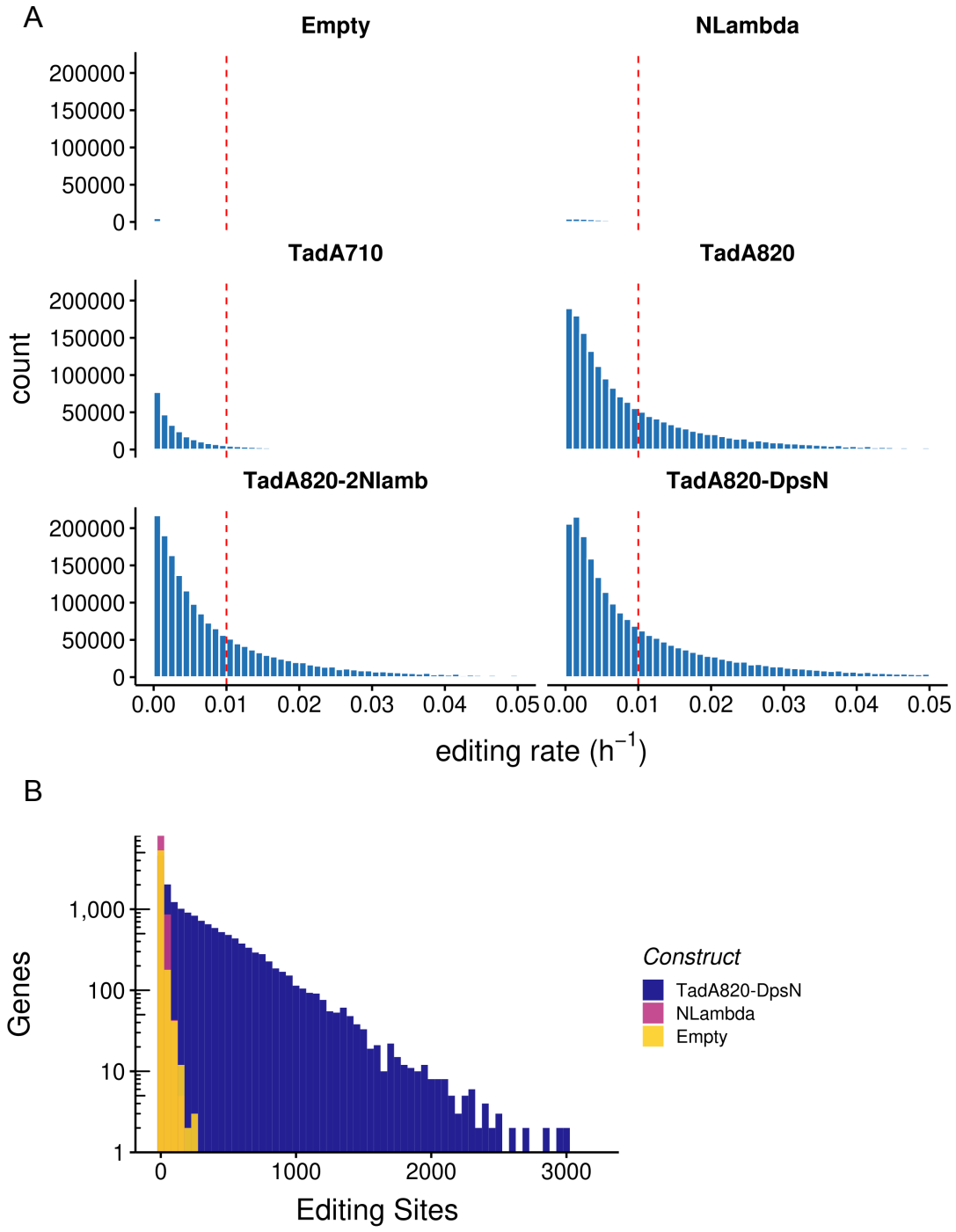


Figure 35 The editing rates and distribution of editing sites of the second-generation editors across the human transcriptome.

(A) A calibration experiment consisting of two timepoints (0 and 8 hours after actinomycin D) was performed. A linear model was fit to the numbers of edits at detected sites to approximate the editing

rate lambda from Equation 1 (Main). Histograms of the mean editing rates are shown (n=3). TadA8.20-DpsN vector courtesy of Karl Brune.

(B) Histogram showing the transcriptome-wide distribution of editing sites identified by JACUSA2 for the empty vector control (endogenous editing), Nlambda-ADAR2 and TadA8.20-Dps-N.

14.4. Future work with TadA editing constructs

Unfortunately, the end of my PhD arrived to quickly for me to be able to use the second-generation editors in any discovery experiments. However, I believe they will be critical for using Timestamps in single cell RNA-seq as previously discussed. Furthermore, these editing constructs will be critical for studying genes which are not well edited in humans or even in using Timestamps in non-primate model organisms, which have low levels of RNA editing compared to primates and humans (Tan et al. 2017; Mansi et al. 2021; Bazak et al. 2014).

14.5. Methods for Chapter 14

For the second generation of editors, I used a chassis based on the E. coli tRNA-specific adenosine deaminase (TadA). Plasmids for ADAR2(E488Q)-Nlambda (the top-performing editor from the first generation), TadA7.10 (Gaudelli et al. 2020), TadA8.20 (Xiao et al. 2023), TadA8.20–Nlamdba and TadA8.20–Dps-N (Kwon and Giessen 2022; Park et al. 2020) (Dps-N construct courtesy of Karl Brune) were transfected into HEK293 cells and the experiment performed as for the first generation editors, except, for the second-generation, I sequenced two time points so that editing rates could be fit. Cell culture, cloning and transfections were performed as in Chapter 11 (11.2).

All plasmids used in this study are listed in Table 8.

15. Discussion

Existing methods to encode temporal information about the cell's transcriptional state have relied on three encoding formats: DNA (e.g. the Cas1-Cas2 integrase system (Shipman et al. 2016), RNA (e.g. SLAMseq, an RNA pulse-chase method), or protein (e.g. Expression Recording Islands (Linghu et al. 2023; Lin et al. 2023). These approaches have limited utility for at least one of three main reasons: (i) inability to multiplex measurements over many genes (ii) low-throughout read-out such as fluorescence microscopy (iii) cannot record the absolute time at which events occur. Notably, SLAMseq allows parallel observation of thousands of genes, but reports only if an mRNA is old or new relative to a single timepoint (the addition of the nucleotide analogue).

RNA Timestamps (Rodriques et al. 2021) overcomes this limitation by encoding the ages of single mRNA molecules as A-to-G transitions. However, this method required the RNA of interest to be fused to an MCP binding motif in the 3'UTR in to recruit a transiently expressed ADAR-MCP fusion protein (Rodriques et al. 2021). The authors were able to infer the age of single mRNA molecules on the scale of hours from only the empirical distributions of RNA edits per molecule. This approach is, however, difficult to scale to hundreds of mRNAs and *in vivo* implementations would require genetically engineered organisms.

I have shown how endogenous A-to-I editing can be used to infer the ages of individual RNA molecules, and to infer transcriptional histories and transcriptional programs in single, unmodified primary human cells. By using an initial calibration experiment, where transcription is paused with actinomycin D, and cells are lysed and known increments, I was able to fit a model of RNA editing to editing sites across the transcriptome. In subsequent discovery experiments, I used the parameters of this model to infer the age of genes (Illumina) or single transcripts (nanopore/PacBio). The ability to obtain single molecule age estimates hinged on the availability of scalable long-read sequencing. Recent improvements in the accuracy of nanopore sequencing were critical in these efforts.

This method can be extended without modification to infer temporal dynamics in virtually any human system that lends itself to RNA sequencing, such as human tissue biopsies. One major advantage is that Timestamps does not require genetic engineering, as I demonstrated in primary human monocytes; a cell type that is notoriously difficult to engineer. In human monocytes, we were able to show that differential age analysis provides orthogonal information to traditional differential expression analysis, because many genes undergo transient changes in expression. We took this analysis further and identified clusters of differentially aging genes whose stability is regulated by common miRNAs. We identified known transcription factor and miRNA regulators of monocyte differentiation and putative new ones which should be the subject of further investigation. We found that the transcription factor ZF-5 may regulate or be regulated by a cluster of co-aging genes. Interestingly, ZF-5 has been critically implicated in monocyte differentiation in Zebrafish but only now in humans.

The principal deficiency of this system is that not all genes (or animals) have sufficient editing to be studied with Timestamps. The temporal resolution and breadth of genes which can be measured by this technique could be improved by incorporating other forms of RNA edits such as methylation (which can be detected using nanopore direct RNA sequencing), or by using an exogenous, engineered RNA editing protein, such as the TadA constructs I created in this work. Although the first generation Nlambda-ADAR was not useful, I was able to create a new class of editors which will be useful in future single cell or non-primate animal studies using Timestamps.

I began to explore the utility of Timestamps in functional brain mapping at the single cell level. Timestamps could be used to distinguish neuronal responses to temporally separated stimuli, thus enabling single cell RNA-seq data to discriminate the responsiveness of neurones to diverse, but temporally separated stimuli. This is important because while RNA-seq can be scaled to take measurements from thousands of cells, it is fundamentally

destructive. I showed that the Nlambda-ADAR editor could be successfully delivered to the mouse superior colliculus and the double positive widefield neurons could be sorted by FACS and sequenced. In the future, one of the TadA editors such as TadA8.20-DpsN could be utilised.

The ability to make continuous measurements of hundreds of genes across the transcriptome composition has until now been elusive. I have shown that endogenous RNA editing by ADAR enzymes forms a continuous and distributed ledger of past transcriptional regulation.

16. References

"A More Personal View of Human-Gene Regulation." 2017. Nature 550 (7675): 157–157. https://doi.org/10.1038/550157a.

Aitken, M. J. 1990. Science-Based Dating in Archaeology. Longman Archaeology Series. London ; New York: Longman.

Alon, Shahar, Sandra C Garrett, Erez Y Levanon, Sara Olson, Brenton R Graveley, Joshua J C Rosenthal, and Eli Eisenberg. 2015. "The Majority of Transcripts in the Squid Nervous System Are Extensively Recoded by A-to-I RNA Editing." Edited by Roderic Guigó. eLife 4 (January):e05198. https://doi.org/10.7554/eLife.05198.

Bahn, Jae Hoon, Jaegyoon Ahn, Xianzhi Lin, Qing Zhang, Jae-Hyung Lee, Mete Civelek, and Xinshu Xiao. 2015. "Genomic Analysis of ADAR1 Binding and Its Involvement in Multiple RNA Processing Pathways." Nature Communications 6 (March):6355. https://doi.org/10.1038/ncomms7355.

Bass, B. L., and H. Weintraub. 1988. "An Unwinding Activity That Covalently Modifies Its Double-Stranded RNA Substrate." Cell 55 (6): 1089–98. https://doi.org/10.1016/0092-8674(88)90253-x.

Bass, Brenda L. 2024. "Adenosine Deaminases That Act on RNA, Then and Now." RNA 30 (5): 521–29. https://doi.org/10.1261/rna.079990.124.

Bazak, Lily, Ami Haviv, Michal Barak, Jasmine Jacob-Hirsch, Patricia Deng, Rui Zhang, Farren J. Isaacs, et al. 2014. "A-to-I RNA Editing Occurs at over a Hundred Million Genomic Sites, Located in a Majority of Human Genes." Genome Research 24 (3): 365–76. https://doi.org/10.1101/gr.164749.113.

Benne, Rob, Janny Van Den Burg, Just P. J. Brakenhoff, Paul Sloof, Jacques H. Van Boom, and Marijke C. Tromp. 1986. "Major Transcript of the Frameshifted Coxll Gene from Trypanosome

Mitochondria Contains Four Nucleotides That Are Not Encoded in the DNA." Cell 46 (6): 819–26. https://doi.org/10.1016/0092-8674(86)90063-2.

Bentley, David R., Shankar Balasubramanian, Harold P. Swerdlow, Geoffrey P. Smith, John Milton, Clive G. Brown, Kevin P. Hall, et al. 2008. "Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry." Nature 456 (7218): 53–59. https://doi.org/10.1038/nature07517.

Bertrand, Edouard, Pascal Chartrand, Matthias Schaefer, Shailesh M. Shenoy, Robert H. Singer, and Roy M. Long. 1998. "Localization of ASH1 mRNA Particles in Living Yeast." Molecular Cell 2 (4): 437–45. https://doi.org/10.1016/S1097-2765(00)80143-4.

Bhalla, Tarun, Joshua J. C. Rosenthal, Miguel Holmgren, and Robert Reenan. 2004. "Control of Human Potassium Channel Inactivation by Editing of a Small mRNA Hairpin." Nature Structural & Molecular Biology 11 (10): 950–56. https://doi.org/10.1038/nsmb825.

Bhat, Pooja, Luis E Cabrera-quio, Veronika A Herzog, Nina Fasching, Andrea Pauli, Stefan L Ameres, Pooja Bhat, et al. 2023. "Resource SLAMseq Resolves the Kinetics of Maternal and Zygotic Gene Expression during Early Zebrafish Embryogenesis Ll Ll SLAMseq Resolves the Kinetics of Maternal and Zygotic Gene Expression during Early Zebrafish Embryogenesis." CellReports 42 (2): 112070. https://doi.org/10.1016/j.celrep.2023.112070.

Bhattacharya, Aniket, Vineet Jha, Khushboo Singhal, Mahar Fatima, Dayanidhi Singh, Gaura Chaturvedi, Dhwani Dholakia, et al. 2021. "Multiple Alu Exonization in 30UTR of a Primate-Specific Isoform of CYP20A1 Creates a Potential miRNA Sponge." Genome Biology and Evolution 13 (1): 1–19. https://doi.org/10.1093/gbe/evaa233.

Binz, Regina L., Erming Tian, Ratan Sadhukhan, Daohong Zhou, Martin Hauer-Jensen, and Rupak Pathak. 2019. "Identification of Novel Breakpoints for Locus- and Region-Specific Translocations in 293 Cells by Molecular Cytogenetics before and after Irradiation." Scientific Reports 9 (July):10554. https://doi.org/10.1038/s41598-019-47002-0.

Birk, Matthew A., Noa Liscovitch-Brauer, Matthew J. Dominguez, Sean McNeme, Yang Yue, J. Damon Hoff, Itamar Twersky, et al. 2023. “Temperature-Dependent RNA Editing in Octopus Extensively Recodes the Neural Proteome.” Cell 186 (12): 2544-2555.e13. https://doi.org/10.1016/j.cell.2023.05.004.

Bogaerts, Bert, An Van Den Bossche, Bavo Verhaegen, Laurence Delbrassinne, Wesley Mattheus, Stéphanie Nouws, Maxime Godfroid, et al. 2024. “Closing the Gap: Oxford Nanopore Technologies R10 Sequencing Allows Comparable Results to Illumina Sequencing for SNP-Based Outbreak Investigation of Bacterial Pathogens.” Edited by Daniel D. Rhoads. Journal of Clinical Microbiology 62 (5): e01576-23. https://doi.org/10.1128/jcm.01576-23.

Boström, K, Z Garcia, K S Poksay, D F Johnson, A J Lusis, and T L Innerarity. 1990. “Apolipoprotein B mRNA Editing. Direct Determination of the Edited Base and Occurrence in Non-Apolipoprotein B-Producing Cell Lines.” Journal of Biological Chemistry 265 (36): 22446–52. https://doi.org/10.1016/S0021-9258(18)45725-0.

Bouska, Alyssa, and Christine M. Eischen. 2009. “Murine Double Minute 2: P53-Independent Roads Lead to Genome Instability or Death.” Trends in Biochemical Sciences 34 (6): 279–86. https://doi.org/10.1016/j.tibs.2009.02.006.

Canard, Bruno, and Robert S. Sarfati. 1994. “DNA Polymerase Fluorescent Substrates with Reversible 3'-Tags.” Gene 148 (1): 1–6. https://doi.org/10.1016/0378-1119(94)90226-7.

Cattenoz, Pierre B., Ryan J. Taft, Eric Westhof, and John S. Mattick. 2013. “Transcriptome-Wide Identification of A > I RNA Editing Sites by Inosine Specific Cleavage.” RNA 19 (2): 257–70. https://doi.org/10.1261/rna.036202.112.

Choi, Junhong, Wei Chen, Anna Minkina, Florence M. Chardon, Chase C. Suiter, Samuel G. Regalado, Silvia Domcke, et al. 2022. “A Time-Resolved, Multi-Symbol Molecular Recorder via Sequential Genome Editing.” Nature 608 (7921): 98–107. https://doi.org/10.1038/s41586-022-04922-8.

Cockroft, Scott L., John Chu, Manuel Amorin, Hagan Bayley, and M. Reza Ghadiri. 2008. “A Single-Molecule Nanopore Device Detects DNA Polymerase Activity With Single-Nucleotide Resolution.” Journal of the American Chemical Society 130 (3): 818–20. https://doi.org/10.1021/ja077082c.

Covello, P. S., and M. W. Gray. 1989. “RNA Editing in Plant Mitochondria.” Nature 341 (6243): 662–66. https://doi.org/10.1038/341662a0.

Cox, David B.T., Jonathan S. Gootenberg, Omar O. Abudayyeh, Brian Franklin, Max J. Kellner, Julia Joung, and Feng Zhang. 2017. “RNA Editing with CRISPR-Cas13.” Science 358 (6366): 1019–27. https://doi.org/10.1126/science.aaq0180.

De Simoni, Anna, and Lily MY Yu. 2006. “Preparation of Organotypic Hippocampal Slice Cultures: Interface Method.” Nature Protocols 1 (3): 1439–45. https://doi.org/10.1038/nprot.2006.228.

Delahaye, Clara, and Jacques Nicolas. 2021. “Sequencing DNA with Nanopores: Troubles and Biases.” PLoS ONE 16 (10): e0257521. https://doi.org/10.1371/journal.pone.0257521.

Dijk, Erwin L. van, Yan Jaszczyszyn, Delphine Naquin, and Claude Thermes. n.d. “The Third Revolution in Sequencing Technology.” Accessed September 18, 2024. https://www.cell.com/trends/genetics/abstract/S0168-9525(18)30096-9.

DRAGER, URSULA C, and DAVID H HUBEL. 1975. “Physiology of Visual Cells in Mouse Superior Colliculus and Correlation with Somatosensory and Auditory Input.” Nature 253 (5488): 203–4. https://doi.org/10.1038/253203a0.

Eid, John, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, et al. 2009. “Real-Time DNA Sequencing from Single Polymerase Molecules.” Science 323 (5910): 133–38. https://doi.org/10.1126/science.1162986.

Fordham, Jezrom B, Afsar R Naqvi, and Salvador Nares. 2015. “Regulation of miR-24, miR-30b, and miR-142-3p during Macrophage and Dendritic Cell Differentiation Potentiates

Innate Immunity." Journal of Leukocyte Biology 98 (2): 195–207. https://doi.org/10.1189/jlb.1a1014-519rr.

Gale, Samuel D., and Gabe J. Murphy. 2014. "Distinct Representation and Distribution of Visual Information by Specific Cell Types in Mouse Superficial Superior Colliculus." Journal of Neuroscience 34 (40): 13458–71. https://doi.org/10.1523/JNEUROSCI.2768-14.2014.

Gao, Zongliang, Elena Herrera-Carrillo, and Ben Berkhout. 2018. "Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III." Molecular Therapy - Nucleic Acids 10 (March): 36–44. https://doi.org/10.1016/j.omtn.2017.11.006.

Gaudelli, Nicole M., Alexis C. Komor, Holly A. Rees, Michael S. Packer, Ahmed H. Badran, David I. Bryson, and David R. Liu. 2017. "Programmable Base Editing of T to G C in Genomic DNA without DNA Cleavage." Nature 551 (7681): 464–71. https://doi.org/10.1038/nature24644.

Gaudelli, Nicole M., Dieter K. Lam, Holly A. Rees, Noris M. Solá-Esteves, Luis A. Barrera, David A. Born, Aaron Edwards, et al. 2020. "Directed Evolution of Adenine Base Editors with Increased Activity and Therapeutic Application." Nature Biotechnology 38 (7): 892–900. https://doi.org/10.1038/s41587-020-0491-6.

Gerfen, Charles R., Ronald Paletzki, and Nathaniel Heintz. 2013. "GENSAT BAC Cre-Recombinase Driver Lines to Study the Functional Organization of Cerebral Cortical and Basal Ganglia Circuits." Neuron 80 (6): 1368–83. https://doi.org/10.1016/j.neuron.2013.10.016.

Gommans, Willemijn M., Sean P. Mullen, and Stefan Maas. 2009. "RNA Editing: A Driving Force for Adaptive Evolution?" BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology 31 (10): 1137–45. https://doi.org/10.1002/bies.200900045.

Graveley, Brenton R., Angela N. Brooks, Joseph W. Carlson, Michael O. Duff, Jane M. Landolin, Li Yang, Carlo G. Artieri, et al. 2011. "The Developmental Transcriptome of Drosophila Melanogaster." Nature 471 (7339): 473–79. https://doi.org/10.1038/nature09715.

Grünewald, Julian, Ronghao Zhou, Sara P. Garcia, Sowmya Iyer, Caleb A. Lareau, Martin J. Aryee, and J. Keith Joung. 2019. “Transcriptome-Wide off-Target RNA Editing Induced by CRISPR-Guided DNA Base Editors.” Nature 569 (7756): 433–37. https://doi.org/10.1038/s41586-019-1161-z.

Grünewald, Julian, Ronghao Zhou, Sowmya Iyer, Caleb A Lareau, Sara P Garcia, Martin J Aryee, and J Keith Joung. 2019. “CRISPR DNA Base Editors with Reduced RNA Off-Target and Self-Editing Activities.” Nature Biotechnology 37 (September). https://doi.org/10.1038/s41587-019-0236-6.

Herzog, Veronika A., Brian Reichholf, Tobias Neumann, Philipp Rescheneder, Pooja Bhat, Thomas R. Burkard, Wiebke Wlotzka, Arndt Von Haeseler, Johannes Zuber, and Stefan L. Ameres. 2017. “Thiol-Linked Alkylation of RNA to Assess Expression Dynamics.” Nature Methods 14 (12): 1198–1204. https://doi.org/10.1038/nmeth.4435.

Hnisz, D, B Abraham, T Lee, A Lau, V Saint-Andre, A Sigova, H Hoke, and R Young. 2014. “Transcriptional Super-Enhancers Connected to Cell Identity and Disease.” Cell 155 (4): 1–24. https://doi.org/10.1016/j.cell.2013.09.053.Transcriptional.

Hoy, Jennifer L., Hannah I. Bishop, and Cristopher M. Niell. 2019. “Defined Cell Types in Superior Colliculus Make Distinct Contributions to Prey Capture Behavior in the Mouse.” Current Biology 29 (23): 4130-4138.e5. https://doi.org/10.1016/j.cub.2019.10.017.

Hume, Richard I., Raymond Dingledine, and Stephen F. Heinemann. 1991. “Identification of a Site in Glutamate Receptor Subunits That Controls Calcium Permeability.” Science 253 (5023): 1028–31. https://doi.org/10.1126/science.1653450.

Illumina. 2015. “An Introduction to Next-Generation Sequencing Technology.” Https://Www.Illumina.Com. https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf#page=2.07.

Jain, Miten, Robin Abu-Shumays, Hugh E. Olsen, and Mark Akeson. 2022. “Advances in Nanopore Direct RNA Sequencing.” Nature Methods 19 (10): 1160–64. https://doi.org/10.1038/s41592-022-01633-w.

Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. “The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community.” Genome Biology 17 (November):239. https://doi.org/10.1186/s13059-016-1103-0.

Jin, Yetao, Shelya X. Zeng, Xiao-Xin Sun, Hunjoo Lee, Christine Blattner, Zhixiong Xiao, and Hua Lu. 2008. “MDMX Promotes Proteasomal Turnover of P21 at G1 and Early S Phases Independently of, but in Cooperation with, MDM2.” Molecular and Cellular Biology 28 (4): 1218–29. https://doi.org/10.1128/MCB.01198-07.

Jonas, Peter, and Nail Burnashev. 1995. “Molecular Mechanisms Controlling Calcium Entry through AMPA-Type Glutamate Receptor Channels.” Neuron 15 (5): 987–90. https://doi.org/10.1016/0896-6273(95)90087-X.

Kalhor, Reza, Prashant Mali, and George M. Church. 2017. “Rapidly Evolving Homing CRISPR Barcodes.” Nature Methods 14 (2): 195–200. https://doi.org/10.1038/nmeth.4108.

Knutson, Steve D., and Jennifer M. Heemstra. 2020. “EndoVIPER-Seq for Improved Detection of A-to-I Editing Sites in Cellular RNA.” Current Protocols in Chemical Biology 12 (2): e82. https://doi.org/10.1002/cpch.82.

Kolberg, Liis, and Uku Raudvere. 2024. “Gprofiler2: Interface to the ‘g:Profiler’ Toolset.” https://cran.r-project.org/web/packages/gprofiler2/index.html.

Kolberg, Liis, Uku Raudvere, Ivan Kuzmin, Priit Adler, Jaak Vilo, and Hedi Peterson. 2023. “G:Profiler-Interoperable Web Service for Functional Enrichment Analysis and Gene Identifier Mapping (2023 Update).” Nucleic Acids Research 51 (W1): W207–12. https://doi.org/10.1093/nar/gkad347.

Kuttan, Ashani, and Brenda L. Bass. 2012. “Mechanistic Insights into Editing-Site Specificity of ADARs.” Proceedings of the National Academy of Sciences of the United States of America 109 (48). https://doi.org/10.1073/pnas.1212548109.

Kwon, Seokmu, and Tobias W. Giessen. 2022. “Engineered Protein Nanocages for Concurrent RNA and Protein Packaging In Vivo.” ACS Synthetic Biology 11 (10): 3504–15. https://doi.org/10.1021/acssynbio.2c00391.

La Manno, Gioele, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, et al. 2018. “RNA Velocity of Single Cells.” Nature 560 (7719): 494–98. https://doi.org/10.1038/s41586-018-0414-6.

Lander, Eric S., Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” Nature 409 (6822): 860–921. https://doi.org/10.1038/35057062.

Lau, P P, S H Chen, J C Wang, and L Chan. 1990. “A 40 Kilodalton Rat Liver Nuclear Protein Binds Specifically to Apolipoprotein B mRNA around the RNA Editing Site.” Nucleic Acids Research 18 (19): 5817–21.

Lawrence, Toby, and Gioacchino Natoli. 2011. “Transcriptional Regulation of Macrophage Polarization: Enabling Diversity with Identity.” Nature Reviews Immunology 11 (11): 750–61. https://doi.org/10.1038/nri3088.

Lazinski, David W., E. Grzadzielska, and A. Das. 1989. “Sequence-Specific Recognition of RNA Hairpins by Bacteriophage Antiterminators Requires a Conserved Arginine-Rich Motif.” PubMed. https://pubmed.ncbi.nlm.nih.gov/2477156/.

Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” Bioinformatics 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, J., J. Reisner, H. Pham, S. Olafsson, and S. Vardeman. 2020. “Biclustering with Missing Data.” Information Sciences 510 (February):304–16. https://doi.org/10.1016/j.ins.2019.09.047.

Li, Jin Billy, Erez Y. Levanon, Jung-Ki Yoon, John Aach, Bin Xie, Emily LeProust, Kun Zhang, Yuan Gao, and George M. Church. 2009. “Genome-Wide Identification of Human RNA Editing Sites by Parallel DNA Capturing and Sequencing.” Science 324 (5931): 1210–13. https://doi.org/10.1126/science.1170995.

Licht, Konstantin, Utkarsh Kapoor, Fabian Amman, Ernesto Picardi, David Martin, Prajakta Bajad, and Michael F. Jantsch. 2019. “A High Resolution A-to-I Editing Map in the Mouse Identifies Editing Events Controlled by Pre-mRNA Splicing.” Genome Research 29 (9): 1453–63. https://doi.org/10.1101/gr.242636.118.

Lin, Dingchang, Xiuyuan Li, Eric Moult, Pojeong Park, Benjamin Tang, Hao Shen, Jonathan B. Grimm, et al. 2023. “Time-Tagged Ticker Tapes for Intracellular Recordings.” Nature Biotechnology. https://doi.org/10.1038/s41587-022-01524-7.

Linghu, Changyang, Bobae An, Monika Shpokayte, Orhan T. Celiker, Nava Shmoel, Ruihan Zhang, Chi Zhang, et al. 2023. “Recording of Cellular Physiological Histories along Optically Readable Self-Assembling Protein Chains.” Nature Biotechnology. https://doi.org/10.1038/s41587-022-01586-7.

Liu, Y., C. X. George, J. B. Patterson, and C. E. Samuel. 1997. “Functionally Distinct Double-Stranded RNA-Binding Domains Associated with Alternative Splice Site Variants of the Interferon-Inducible Double-Stranded RNA-Specific Adenosine Deaminase.” The Journal of Biological Chemistry 272 (7): 4419–28. https://doi.org/10.1074/jbc.272.7.4419.

Los, Georgyi V, Lance P Encell, Mark G McDougall, Danette D Hartzell, Natasha Karassina, Chad Zimprich, Monika G Wood, et al. 2008. “HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis.” ACS Chemical Biology 3 (6): 373–82. https://doi.org/10.1021/cb800025k.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” Genome Biology 15 (12): 1–21. https://doi.org/10.1186/s13059-014-0550-8.

Loveless, Theresa B., Joseph H. Grotts, Mason W. Schechter, Elmira Forouzmand, Courtney K. Carlson, Bijan S. Agahi, Guohao Liang, et al. 2021. “Lineage Tracing and Analog Recording in Mammalian Cells by Single-Site DNA Writing.” Nature Chemical Biology 17 (6): 739–47. https://doi.org/10.1038/s41589-021-00769-8.

Lu, Daniel, Tracy Yamawaki, Hong Zhou, Wen-Yu Chou, Mark Chhoa, Edwin Lamas, Sabine S Escobar, et al. 2019. “Limited Differential Expression of miRNAs and Other Small RNAs in LPS-Stimulated Human Monocytes.” PloS One 14 (3): e0214296. https://dx.doi.org/10.1371/journal.pone.0214296.

Maeß, Marten B., Stefanie Sendelbach, and Stefan Lorkowski. 2010. “Selection of Reliable Reference Genes during THP-1 Monocyte Differentiation into Macrophages.” BMC Molecular Biology 11:1–8. https://doi.org/10.1186/1471-2199-11-90.

Maitra, Raj D., Jungsuk Kim, and William B. Dunbar. 2012. “Recent Advances in Nanopore Sequencing.” Electrophoresis 33 (23): 3418–28. https://doi.org/10.1002/elps.201200272.

Mansi, Luigi, Marco Antonio Tangaro, Claudio Lo Giudice, Tiziano Flati, Eli Kopel, Amos Avraham Schaffer, Tiziana Castrignano, Giovanni Chillemi, Graziano Pesole, and Ernesto Picardi. 2021. “REDIportal: Millions of Novel A-to-I RNA Editing Events from Thousands of RNAseq Experiments.” Nucleic Acids Research 49 (D1): D1012–19. https://dx.doi.org/10.1093/nar/gkaa916.

Matsushima, Wayo, Veronika A Herzog, Tobias Neumann, Katharina Gapp, Johannes Zuber, Stefan L Ameres, and Eric A Miska. 2018. “SLAM-ITseq: Sequencing Cell Type-Specific Transcriptomes without Cell Sorting.” Development 145 (13): dev164640. https://doi.org/10.1242/dev.164640.

Matthews, Melissa M., Justin M. Thomas, Yuxuan Zheng, Kiet Tran, Kelly J. Phelps, Anna I. Scott, Jocelyn Havel, Andrew J. Fisher, and Peter A. Beal. 2016. “Structures of Human ADAR2 Bound to dsRNA Reveal Base-Flipping Mechanism and Basis for Site Selectivity.” Nature Structural and Molecular Biology 23 (5): 426–33. https://doi.org/10.1038/nsmb.3203.

Meran, Laween, Isobel Massie, Sara Campinoti, Anne E. Weston, Riana Gaifulina, Lucinda Tullie, Peter Faull, et al. 2020. “Engineering Transplantable Jejunal Mucosal Grafts Using Patient-Derived Organoids from Children with Intestinal Failure.” Nature Medicine 26 (10): 1593–1601. https://doi.org/10.1038/s41591-020-1024-z.

Meryet-Figuiere, Matthieu, Babak Alaei-Mahabadi, Mohamad Moustafa Ali, Sanhita Mitra, Santhilal Subhash, Gaurav Kumar Pandey, Erik Larsson, and Chandrasekhar Kanduri. 2014. “Temporal Separation of Replication and Transcription during S-Phase Progression.” Cell Cycle 13 (20): 3241–48. https://doi.org/10.4161/15384101.2014.953876.

Montiel-Gonźalez, Maria Fernanda, Isabel C. Vallecillo-Viejo, and Joshua J.C. Rosenthal. 2016. “An Efficient System for Selectively Altering Genetic Information within mRNAs.” Nucleic Acids Research 44 (21): 1–12. https://doi.org/10.1093/nar/gkw738.

Muhar, Matthias, Anja Ebert, Tobias Neumann, Christian Umkehrer, Julian Jude, Corinna Wieshofer, Philipp Rescheneder, et al. 2018. “SLAM-Seq Defines Direct Gene-Regulatory Functions of the BRD4-MYC Axis.” Science 360 (6390): 800–805. https://doi.org/10.1126/science.aao2793.

Naqvi, Afsar Raza, Jezrom B. Fordham, and Salvador Nares. 2015. “miR-24, miR-30b, and miR-142-3p Regulate Phagocytosis in Myeloid Inflammatory Cells.” The Journal of Immunology 194 (4): 1916–27. https://doi.org/10.4049/jimmunol.1401893.

Neeman, Yossef, Erez Y. Levanon, Michael F. Jantsch, and Eli Eisenberg. 2006. “RNA Editing Level in the Mouse Is Determined by the Genomic Repeat Repertoire.” RNA 12 (10): 1802–9. https://doi.org/10.1261/rna.165106.

Ni, Ying, Xudong Liu, Zemenu Mengistie Simeneh, Mengsu Yang, and Runsheng Li. 2023. “Benchmarking of Nanopore R10.4 and R9.4.1 Flow Cells in Single-Cell Whole-Genome Amplification and Whole-Genome Shotgun Sequencing.” Computational and Structural Biotechnology Journal 21:2352–64. https://doi.org/10.1016/j.csbj.2023.03.038.

Niswender, Colleen M., Sara C. Copeland, Katharine Herrick-Davis, Ronald B. Emeson, and Elaine Sanders-Bush. 1999. “RNA Editing of the Human Serotonin 5-Hydroxytryptamine 2C Receptor Silences Constitutive Activity*.” Journal of Biological Chemistry 274 (14): 9472–78. https://doi.org/10.1074/jbc.274.14.9472.

Park, Chan, Yoontae Jin, Young Jun Kim, Hotcherl Jeong, and Baik L. Seong. 2020. “RNA-Binding as Chaperones of DNA Binding Proteins from Starved Cells.” Biochemical and Biophysical Research Communications 524 (2): 484–89. https://doi.org/10.1016/j.bbrc.2020.01.121.

Passmore, Lori A., and Jeff Coller. 2022. “Roles of mRNA Poly(A) Tails in Regulation of Eukaryotic Gene Expression.” Nature Reviews Molecular Cell Biology 23 (2): 93–106. https://doi.org/10.1038/s41580-021-00417-y.

Patterson, John B., and Charles E. Samuel. 1995. “Expression and Regulation by Interferon of a Double-Stranded-RNA-Specific Adenosine Deaminase from Human Cells: Evidence for Two Forms of the Deaminase.” Molecular and Cellular Biology 15 (10): 5376–88. https://doi.org/10.1128/MCB.15.10.5376.

Paz-Yaacov, Nurit, Erez Y. Levanon, Eviatar Nevo, Yaron Kinar, Alon Harmelin, Jasmine Jacob-Hirsch, Ninette Amariglio, Eli Eisenberg, and Gideon Rechavi. 2010. “Adenosine-to-Inosine RNA Editing Shapes Transcriptome Diversity in Primates.” Proceedings of the National Academy of Sciences 107 (27): 12174–79. https://doi.org/10.1073/pnas.1006183107.

Pestal, Kathleen, Cory C. Funk, Jessica M. Snyder, Nathan D. Price, Piper M. Treuting, and Daniel B. Stetson. 2015. “Isoforms of RNA-Editing Enzyme ADAR1 Independently Control

Nucleic Acid Sensor MDA5-Driven Autoimmunity and Multi-Organ Development." Immunity 43 (5): 933–44. https://doi.org/10.1016/j.immuni.2015.11.001.

Piechotta, Michael, Isabel S. Naarmann-de Vries, Qi Wang, Janine Altmüller, and Christoph Dieterich. 2022. "RNA Modification Mapping with JACUSA2." Genome Biology 23 (1): 1–11. https://doi.org/10.1186/s13059-022-02676-0.

Piechotta, Michael, Emanuel Wyler, Uwe Ohler, Markus Landthaler, and Christoph Dieterich. 2017. "JACUSA: Site-Specific Identification of RNA Editing Events from Replicate Sequencing Data." BMC Bioinformatics 18 (1): 1–15. https://doi.org/10.1186/s12859-016-1432-8.

Pinto, Yishay, Haim Y Cohen, and Erez Y Levanon. 2014. "Mammalian Conserved ADAR Targets Comprise Only a Small Fragment of the Human Editosome." Genome Biology 15 (1): R5. https://doi.org/10.1186/gb-2014-15-1-r5.

Pinto, Yishay, and Erez Y. Levanon. 2019. "Computational Approaches for Detection and Quantification of A-to-I RNA-Editing." Methods, Mining the Epitranscriptome: Detection of RNA editing and RNA modifications, 156 (March):25–31. https://doi.org/10.1016/j.ymeth.2018.11.011.

Polson, A G, and B L Bass. 1994. "Preferential Selection of Adenosines for Modification by Double-Stranded RNA Adenosine Deaminase." The EMBO Journal 13 (23): 5701–11.

Ramaswami, Gokul, Rui Zhang, Robert Piskol, Liam P. Keegan, Patricia Deng, Mary A. O'Connell, and Jin Billy Li. 2013. "Identifying RNA Editing Sites Using RNA Sequencing Data Alone." Nature Methods 10 (2): 128–32. https://doi.org/10.1038/nmeth.2330.

Ravi, Vidhya M, Kevin Joseph, Julian Wurm, Simon Behringer, Nicklas Garrelfs, Paolo d'Errico, Yashar Naseri, et al. 2019. "Human Organotypic Brain Slice Culture: A Novel Framework for Environmental Research in Neuro-Oncology." Life Science Alliance 2 (4): e201900305. https://doi.org/10.26508/lsa.201900305.

Reed, Kathleen S.M., Eric S. Davis, Marielle L. Bond, Alan Cabrera, Eliza Thulson, Ivana Yoseli Quiroga, Shannon Cassel, et al. 2022. “Temporal Analysis Suggests a Reciprocal Relationship between 3D Chromatin Structure and Transcription.” Cell Reports 41 (5): 111567. https://doi.org/10.1016/j.celrep.2022.111567.

Reisner, John, Hieu Pham, and Jing Li. 2021. “Biclustermd: Biclustering with Missing Data.” https://cran.r-project.org/web/packages/biclustermd/index.html.

Rodriques, Samuel G., Linlin M. Chen, Sophia Liu, Ellen D. Zhong, Joseph R. Scherrer, Edward S. Boyden, and Fei Chen. 2021. “RNA Timestamps Identify the Age of Single Molecules in RNA Sequencing.” Nature Biotechnology 39 (3): 320–25. https://doi.org/10.1038/s41587-020-0704-z.

Roquet, Nathaniel, Ava P. Soleimany, Alyssa C. Ferris, Scott Aaronson, and Timothy K. Lu. 2016. “Synthetic Recombinase-Based State Machines in Living Cells.” Science 353 (6297). https://doi.org/10.1126/science.aad8559.

Rosenberg, Brad R., Claire E. Hamilton, Michael M. Mwangi, Scott Dewell, and F. Nina Papavasiliou. 2011. “Transcriptome-Wide Sequencing Reveals Numerous APOBEC1 mRNA-Editing Targets in Transcript 3' UTRs.” Nature Structural & Molecular Biology 18 (2): 230–36. https://doi.org/10.1038/nsmb.1975.

Sakurai, Masayuki, Hiroki Ueda, Takanori Yano, Shunpei Okada, Hideki Terajima, Toutai Mitsuyama, Atsushi Toyoda, Asao Fujiyama, Hitomi Kawabata, and Tsutomu Suzuki. 2014. “A Biochemical Landscape of A-to-I RNA Editing in the Human Brain Transcriptome.” Genome Research 24 (3): 522–34. https://doi.org/10.1101/gr.162537.113.

Salter, Jason D., Ryan P. Bennett, and Harold C. Smith. 2016. “The APOBEC Protein Family: United by Structure, Divergent in Function.” Trends in Biochemical Sciences 41 (7): 578–94. https://doi.org/10.1016/j.tibs.2016.05.001.

Samuel, Charles E. 2019. “Adenosine Deaminase Acting on RNA (ADAR1), a Suppressor of Double-Stranded RNA–Triggered Innate Immune Responses.” Journal of Biological Chemistry 294 (5): 1710–20. https://doi.org/10.1074/jbc.TM118.004166.

Sato, Toshiro, Daniel E. Stange, Marc Ferrante, Robert G. J. Vries, Johan H. van Es, Stieneke van den Brink, Winan J. van Houdt, et al. 2011. “Long-Term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett’s Epithelium.” Gastroenterology 141 (5): 1762–72. https://doi.org/10.1053/j.gastro.2011.07.050.

Schaffer, Amos A, and Erez Y Levanon. 2021. “ALU A-to-I RNA Editing: Millions of Sites and Many Open Questions.” Methods in Molecular Biology (Clifton, N.J.) 2181:149–62. https://dx.doi.org/10.1007/978-1-0716-0787-9_9.

Shendure, Jay, Shankar Balasubramanian, George M. Church, Walter Gilbert, Jane Rogers, Jeffery A. Schloss, and Robert H. Waterston. 2017. “DNA Sequencing at 40: Past, Present and Future.” Nature 550 (7676): 345–53. https://doi.org/10.1038/nature24286.

Shendure, Jay, and Hanlee Ji. 2008. “Next-Generation DNA Sequencing.” Nature Biotechnology 26 (10): 1135–45. https://doi.org/10.1038/nbt1486.

Sheth, Ravi U., and Harris H. Wang. 2018. “DNA-Based Memory Devices for Recording Cellular Events.” Nature Reviews Genetics 19 (11): 718–32. https://doi.org/10.1038/s41576-018-0052-8.

Shi, Yichen, Peter Kirwan, and Frederick J. Livesey. 2012. “Directed Differentiation of Human Pluripotent Stem Cells to Cerebral Cortex Neurons and Neural Networks.” Nature Protocols 7 (10): 1836–46. https://doi.org/10.1038/nprot.2012.116.

Shipman, Seth L., Jeff Nivala, Jeffrey D. Macklis, and George M. Church. 2016. “Molecular Recordings by Directed CRISPR Spacer Acquisition.” Science 353 (6298). https://doi.org/10.1126/science.aaf1175.

Singhal, Khushboo, Sonam Dhamija, and Mitali Mukerji. 2023. “Exonized Alu Repeats in the 3’UTR of a CYP20A1_Alu-LT Transcript Act as a miRNA Sponge.” BMC Research Notes 16 (1): 1–7. https://doi.org/10.1186/s13104-023-06289-z.

Stoddart, David, Andrew J. Heron, Ellina Mikhailova, Giovanni Maglia, and Hagan Bayley. 2009. “Single-Nucleotide Discrimination in Immobilized DNA Oligonucleotides with a Biological Nanopore.” Proceedings of the National Academy of Sciences of the United States of America 106 (19): 7702–7. https://doi.org/10.1073/pnas.0901054106.

Sun, Jianfeng, Martin Philpott, Danson Loi, Shuang Li, Pablo Monteagudo-Mesas, Gabriela Hoffman, Jonathan Robson, et al. 2023. “Correcting PCR Amplification Errors in Unique Molecular Identifiers to Generate Absolute Numbers of Sequencing Molecules.” bioRxiv, January, 2023.04.06.535911. https://doi.org/10.1101/2023.04.06.535911.

———. 2024. “Correcting PCR Amplification Errors in Unique Molecular Identifiers to Generate Accurate Numbers of Sequencing Molecules.” Nature Methods 21 (3): 401–5. https://doi.org/10.1038/s41592-024-02168-y.

Sun, Tao, Qin Li, Jonathan M. Geisinger, Shi-Bin Hu, Boming Fan, Shichen Su, Waitang Tsui, Hongchao Guo, Jinbiao Ma, and Jin Billy Li. 2022. “A Small Subset of Cytosolic dsRNAs Must Be Edited by ADAR1 to Evade MDA5-Mediated Autoimmunity.” bioRxiv. https://doi.org/10.1101/2022.08.29.505707.

Sun, Tony, Yingpu Yu, Xianfang Wu, Ashley Acevedo, Ji Dung Luo, Jiayi Wang, William M. Schneider, et al. 2021. “Decoupling Expression and Editing Preferences of ADAR1 P150 and P110 Isoforms.” Proceedings of the National Academy of Sciences of the United States of America 118 (12): 1–8. https://doi.org/10.1073/pnas.2021757118.

Swift, Joseph, and Gloria Coruzzi. 2017. “A Matter of Time - How Transient Transcription Factor Interactions Create Dynamic Gene Regulatory Networks.” Biochimica et Biophysica Acta 1860 (1): 75–83. https://doi.org/10.1016/j.bbagrm.2016.08.007.

Tan, Meng How, Qin Li, Raghuvaran Shanmugam, Robert Piskol, Jennefer Kohler, Amy N Young, Kaiwen Ivy Liu, et al. 2017. “Dynamic Landscape and Regulation of RNA Editing in Mammals.” Edited by Ardlie K G Aguet F Cummings BB, Gelfand ET, Getz G, Hadley K, Handsaker RE, Huang KH, Kashin S, Karczewski KJ, Lek M, Li X, MacArthur DG, Nedzel JL, Nguyen DT, Noble MS, Segre AV, Trowbridge CA, Tukiainen T, Abell NS, Balliu B, Barshir R, Basha O, Battle A, Bogu GK. Nature, Comment in: Nature. 2017 Oct 11;550(7675):190-191; PMID: 29022577 [https://www.ncbi.nlm.nih.gov/pubmed/29022577]Comment in: Nature. 2017 Oct 11;550(7675):157; PMID: 29022932 [https://www.ncbi.nlm.nih.gov/pubmed/29022932]Comment in: Nature. 2017 Oct 11;550, 550 (7675): 249–54. https://dx.doi.org/10.1038/nature24041.

Tirosh, Itay, Benjamin Izar, Sanjay M. Prakadan, Marc H. Wadsworth, Daniel Treacy, John J. Trombetta, Asaf Rotem, et al. 2016. “Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq.” Science 352 (6282): 189–96. https://doi.org/10.1126/science.aad0501.

Tyner, Jeffrey W., Cristina E. Tognon, Daniel Bottomly, Beth Wilmot, Stephen E. Kurtz, Samantha L. Savage, Nicola Long, et al. 2018. “Functional Genomic Landscape of Acute Myeloid Leukaemia.” Nature 562 (7728): 526–31. https://doi.org/10.1038/s41586-018-0623-z.

Vogel, Paul, Matin Moschref, Qin Li, Tobias Merkle, Karthika D. Selvasaravanan, Jin Billy Li, and Thorsten Stafforst. 2018. “Efficient and Precise Editing of Endogenous Transcripts with SNAP-Tagged ADARs.” Nature Methods 15 (7): 535–38. https://doi.org/10.1038/s41592-018-0017-z.

Wang, Xiliang, Yao He, Qiming Zhang, Xianwen Ren, and Zemin Zhang. 2021. “Direct Comparative Analyses of 10X Genomics Chromium and Smart-Seq2.” Genomics, Proteomics & Bioinformatics 19 (2): 253–66. https://doi.org/10.1016/j.gpb.2020.02.005.

Wang, Yunhao, Yue Zhao, Audrey Bollas, Yuru Wang, and Kin Fai Au. 2021. “Nanopore Sequencing Technology, Bioinformatics and Applications.” Nature Biotechnology 39 (11): 1348–65. https://doi.org/10.1038/s41587-021-01108-x.

Wang, Yuru, SeHee Park, and Peter A. Beal. 2018. “Selective Recognition of RNA Substrates by ADAR Deaminase Domains.” Biochemistry 57 (10): 1640–51. https://doi.org/10.1021/acs.biochem.7b01100.

Wenger, Aaron M., Paul Peluso, William J. Rowell, Pi Chuan Chang, Richard J. Hall, Gregory T. Concepcion, Jana Ebler, et al. 2019. “Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome.” Nature Biotechnology 37 (10): 1155–62. https://doi.org/10.1038/s41587-019-0217-9.

Wong, Swee Kee, Shuji Sato, and David W. Lazinski. 2001. “Substrate Recognition by ADAR1 and ADAR2.” Rna 7 (6): 846–58. https://doi.org/10.1017/S135583820101007X.

Wu, Ye Emily, Lin Pan, Yanning Zuo, Xinmin Li, and Weizhe Hong. 2017. “Detecting Activated Cell Populations Using Single-Cell RNA-Seq.” Neuron 96 (2): 313-329.e6. https://doi.org/10.1016/j.neuron.2017.09.026.

Xiao, Yu Lan, Shun Liu, Ruiqi Ge, Yuan Wu, Chuan He, Mengjie Chen, and Weixin Tang. 2023. “Transcriptome-Wide Profiling and Quantification of N 6-Methyladenosine by Enzyme-Assisted Adenosine Deamination.” Nature Biotechnology. https://doi.org/10.1038/s41587-022-01587-6.

Xiao, Yu-Lan, Shun Liu, Ruiqi Ge, Yuan Wu, Chuan He, Mengjie Chen, and Weixin Tang. 2023. “Transcriptome-Wide Profiling and Quantification of N6-Methyladenosine by Enzyme-Assisted Adenosine Deamination.” Nature Biotechnology 41 (7): 993–1003. https://doi.org/10.1038/s41587-022-01587-6.

Yablonovitch, Arielle L., Patricia Deng, Dionna Jacobson, and Jin Billy Li. 2017. “The Evolution and Adaptation of A-to-I RNA Editing.” PLOS Genetics 13 (11): e1007064. https://doi.org/10.1371/journal.pgen.1007064.

Yu, Xinyang, and Michael J. Buck. 2020. “Pioneer Factors and Their in Vitro Identification Methods.” Molecular Genetics and Genomics : MGG 295 (4): 825–35. https://doi.org/10.1007/s00438-020-01675-9.

Zappia, Luke, and Fabian J Theis. 2021. “Over 1000 Tools Reveal Trends in the Single-Cell RNA-Seq Analysis Landscape.” Genome Biology 22 (1): 301. https://dx.doi.org/10.1186/s13059-021-02519-4.

Zaret, Kenneth S., and Jason S. Carroll. 2011. “Pioneer Transcription Factors: Establishing Competence for Gene Expression.” Genes & Development 25 (21): 2227–41. https://doi.org/10.1101/gad.176826.111.

Zeng, Hongkui, and Joshua R. Sanes. 2017. “Neuronal Cell-Type Classification: Challenges, Opportunities and the Path Forward.” Nature Reviews Neuroscience 18 (9): 530–46. https://doi.org/10.1038/nrn.2017.85.

Zhang, Lin, Gan Wang, Shiwen Chen, Jun Ding, Shiming Ju, Heli Cao, and Hengli Tian. 2016. “Depletion of Thymopoietin Inhibits Proliferation and Induces Cell Cycle Arrest/Apoptosis in Glioblastoma Cells.” World Journal of Surgical Oncology 14 (October):267. https://doi.org/10.1186/s12957-016-1018-y.

17. Appendix

Table 4 Fit data for a subset of 500 randomly chosen editing sites from the calibration data

The original data contains >2E5 editing sites, so a subset of these have been chosen for display. These sites are from the 3' UTRs and lambdas > 0.04. This calibration data set contains editing sites from both the HEK293 Nlambda ADAR calibration and the Cortex calibration and was used in the Monocyte discovery experiments. The fitted values λ and α for Equation 1 are shown.

Experiment	Site UID	Gene Name	Gene ID	λ	alpha	R2	Chromosome	Type
HEK nlambda ADAR	6_111272076_1	MFSD4B	ENSG00000173214	0.0401	0.9718	0.7532	6	three_prime_utr
HEK nlambda ADAR	17_39760300_-1	IKZF3	ENSG00000161405	0.1733	0.7778	0.7149	17	three_prime_utr
HEK nlambda ADAR	5_143225352_1	ARHGAP26	ENSG00000145819	0.1808	0.4078	0.8299	5	three_prime_utr
HEK nlambda ADAR	16_48354346_1	LONP2	ENSG00000102910	0.0575	0.6812	0.7683	16	three_prime_utr
HEK nlambda ADAR	5_157756903_1	LSM11	ENSG00000155858	0.0471	0.871	0.6516	5	three_prime_utr
HEK nlambda ADAR	19_45549973_-1	OPA3	ENSG00000125741	0.1863	0.4363	0.9382	19	three_prime_utr
HEK nlambda ADAR	2_201286638_1	CASP8	ENSG00000064012	0.0785	0.7384	0.9552	2	three_prime_utr
HEK nlambda ADAR	16_53694648_-1	RPGRIP1L	ENSG00000103494	0.1339	0.6066	0.9061	16	three_prime_utr
HEK nlambda ADAR	13_45030893_1	GPALPP1	ENSG00000133114	0.2872	0.596	0.9596	13	three_prime_utr
HEK nlambda ADAR	17_39759243_-1	IKZF3	ENSG00000161405	0.1193	0.9145	0.8817	17	three_prime_utr
HEK nlambda ADAR	17_10677992_-1	SCO1	ENSG00000133028	0.0707	0.7625	0.6529	17	three_prime_utr
HEK nlambda ADAR	1_45512301_1	MMACHC	ENSG00000132763	0.1483	0.4772	0.5106	1	three_prime_utr

HEK nlambda ADAR	19_7085383_1	ZNF557	ENSG00000130544	0.0661	0.9222	0.8119	19	three_prime_utr
HEK nlambda ADAR	2_20676229_1	GDF7	ENSG00000143869	0.0577	0.8262	0.4567	2	three_prime_utr
HEK nlambda ADAR	1_9102301_-1	GPR157	ENSG00000180758	0.0714	0.8868	0.7947	1	three_prime_utr
HEK nlambda ADAR	1_6524850_-1	NOL9	ENSG00000162408	0.0974	0.8102	0.9251	1	three_prime_utr
HEK nlambda ADAR	1_220058616_-1	BPNT1	ENSG00000162813	0.1295	0.1267	0.9215	1	three_prime_utr
HEK nlambda ADAR	12_116851537_1	RNFT2	ENSG00000135119	0.0849	0.946	0.7652	12	three_prime_utr
HEK nlambda ADAR	17_40283455_1	WIPF2	ENSG00000171475	0.1501	0.6242	0.8082	17	three_prime_utr
HEK nlambda ADAR	17_76713301_-1	JMJD6	ENSG00000070495	0.0536	1.0325	0.825	17	three_prime_utr
HEK nlambda ADAR	1_35602629_-1	PSMB2	ENSG00000126067	0.1019	0.5945	0.9952	1	three_prime_utr
HEK nlambda ADAR	17_49829994_1	KAT7	ENSG00000136504	0.0568	0.9377	0.7105	17	three_prime_utr
HEK nlambda ADAR	2_203305304_1	CYP20A1	ENSG00000119004	0.0538	0.8929	0.798	2	three_prime_utr
HEK nlambda ADAR	17_44801118_-1	GJC1	ENSG00000182963	0.4277	0.121	0.9901	17	three_prime_utr
HEK nlambda ADAR	11_72291853_-1	CLPB	ENSG00000162129	0.1341	0.1	0.804	11	three_prime_utr
HEK nlambda ADAR	2_182789321_1	DNAJC10	ENSG00000077232	0.0564	0.8063	0.7078	2	three_prime_utr
HEK nlambda ADAR	2_206118975_-1	NDUFS1	ENSG00000023228	0.0405	0.9731	0.7372	2	three_prime_utr
HEK nlambda ADAR	1_228180775_1	IBA57	ENSG00000181873	0.1689	0.5967	0.6817	1	three_prime_utr
HEK nlambda ADAR	6_111230225_1	SLC16A10	ENSG00000112394	0.4163	0.553	0.9901	6	three_prime_utr
HEK nlambda ADAR	4_70816663_-1	GRSF1	ENSG00000132463	0.1943	0.6345	0.949	4	three_prime_utr

HEK nlambda ADAR	2_206121631_-1	NDUFS1	ENSG00000023228	0.0981	0.6286	0.9159	2	three_prime_utr
HEK nlambda ADAR	19_23222355_-1	ZNF724	ENSG00000196081	0.057	0.9704	0.8837	19	three_prime_utr
HEK nlambda ADAR	5_69282095_-1	CCDC125	ENSG00000183323	0.0686	0.8412	0.9207	5	three_prime_utr
HEK nlambda ADAR	19_39491265_1	TIMM50	ENSG00000105197	0.0457	0.8493	0.8281	19	three_prime_utr
HEK nlambda ADAR	9_128306934_-1	TRUB2	ENSG00000167112	0.2873	0.0853	0.7391	9	three_prime_utr
HEK nlambda ADAR	6_89255353_-1	GABRR2	ENSG00000111886	0.0941	0.7063	0.6978	6	three_prime_utr
HEK nlambda ADAR	1_9101082_-1	GPR157	ENSG00000180758	0.132	0.6484	0.9074	1	three_prime_utr
HEK nlambda ADAR	16_75474964_-1	CHST6	ENSG00000183196	0.0568	0.9994	0.8211	16	three_prime_utr
HEK nlambda ADAR	16_48354307_1	LONP2	ENSG00000102910	0.0823	0.8235	0.8878	16	three_prime_utr
HEK nlambda ADAR	10_119573191_-1	TIAL1	ENSG00000151923	0.083	0.8169	0.856	10	three_prime_utr
HEK nlambda ADAR	9_128308288_-1	TRUB2	ENSG00000167112	0.1144	0.5574	0.7663	9	three_prime_utr
HEK nlambda ADAR	17_62452844_1	METTL2A	ENSG00000087995	0.0802	0.6687	0.8016	17	three_prime_utr
HEK nlambda ADAR	1_204555706_1	MDM4	ENSG00000198625	0.0431	0.8671	0.6949	1	three_prime_utr
HEK nlambda ADAR	6_111230170_1	SLC16A10	ENSG00000112394	0.4755	0.3418	0.9663	6	three_prime_utr
HEK nlambda ADAR	11_88322521_-1	CTSC	ENSG00000109861	0.0455	0.6023	0.4584	11	three_prime_utr
HEK nlambda ADAR	17_43029728_1	RND2	ENSG00000108830	0.3527	0.0211	0.8199	17	three_prime_utr
HEK nlambda ADAR	2_201287277_1	CASP8	ENSG00000064012	0.0413	0.9773	0.9556	2	three_prime_utr
HEK nlambda ADAR	11_111728453_1	SIK2	ENSG00000170145	0.2017	0.3941	0.6959	11	three_prime_utr

HEK nlambda ADAR	19_37633445_-1	ZFP30	ENSG00000120784	0.5484	0.1472	0.8264	19	three_prime_utr
HEK nlambda ADAR	11_68756035_-1	CPT1A	ENSG00000110090	0.0488	0.9003	0.9356	11	three_prime_utr
HEK nlambda ADAR	1_32605544_1	ZBTB8A	ENSG00000160062	0.0408	0.8797	0.7283	1	three_prime_utr
HEK nlambda ADAR	20_3869129_1	MAVS	ENSG00000088888	0.0647	0.2389	0.5439	20	three_prime_utr
HEK nlambda ADAR	17_62451209_1	METTL2A	ENSG00000087995	0.0757	0.8549	0.8082	17	three_prime_utr
HEK nlambda ADAR	1_114584547_-1	DENND2C	ENSG00000175984	0.0541	0.9864	0.8252	1	three_prime_utr
HEK nlambda ADAR	20_33705153_-1	PXMP4	ENSG00000101417	0.0901	0.786	0.7379	20	three_prime_utr
HEK nlambda ADAR	22_17113325_1	IL17RA	ENSG00000177663	0.0839	0.6407	0.7673	22	three_prime_utr
HEK nlambda ADAR	10_72234024_1	ANAPC16	ENSG00000166295	0.1465	0.37	0.8156	10	three_prime_utr
HEK nlambda ADAR	22_40927715_1	XPNPEP3	ENSG00000196236	0.0866	0.8358	0.8562	22	three_prime_utr
HEK nlambda ADAR	13_52694708_1	SUGT1	ENSG00000165416	0.0444	0.9049	0.6558	13	three_prime_utr
HEK nlambda ADAR	4_17801938_-1	DCAF16	ENSG00000163257	0.0862	0.5953	0.9368	4	three_prime_utr
HEK nlambda ADAR	7_4770367_1	FOXK1	ENSG00000164916	0.0803	0.9453	0.8212	7	three_prime_utr
HEK nlambda ADAR	12_71794153_1	RAB21	ENSG00000080371	0.1645	0.7314	0.8721	12	three_prime_utr
HEK nlambda ADAR	7_5971946_-1	PMS2	ENSG00000122512	0.0524	0.7299	0.8498	7	three_prime_utr
HEK nlambda ADAR	17_39257147_-1	FBXL20	ENSG00000108306	0.0744	0.5919	0.7055	17	three_prime_utr
HEK nlambda ADAR	9_125221800_1	RABEPK	ENSG00000136933	0.0456	0.7987	0.488	9	three_prime_utr
HEK nlambda ADAR	22_24573020_1	SNRPD3	ENSG00000100028	0.1414	0.3355	0.945	22	three_prime_utr

HEK nlambda ADAR	14_58550938_1	KIAA0586	ENSG00000100578	0.0455	0.8354	0.7552	14	three_prime_utr
HEK nlambda ADAR	5_37291445_-1	NUP155	ENSG00000113569	0.0724	0.5648	0.8809	5	three_prime_utr
HEK nlambda ADAR	1_32603194_1	ZBTB8A	ENSG00000160062	0.0441	1.0027	0.9082	1	three_prime_utr
HEK nlambda ADAR	16_66887766_1	PDP2	ENSG00000172840	0.0443	0.9771	0.7309	16	three_prime_utr
HEK nlambda ADAR	15_50704179_-1	SPPL2A	ENSG00000138600	0.077	0.738	0.824	15	three_prime_utr
HEK nlambda ADAR	21_42911065_1	NDUFV3	ENSG00000160194	0.0509	0.76	0.5609	21	three_prime_utr
HEK nlambda ADAR	17_44799810_-1	GJC1	ENSG00000182963	0.0559	0.9956	0.8416	17	three_prime_utr
HEK nlambda ADAR	17_75889857_-1	TRIM65	ENSG00000141569	0.137	0.8492	0.9805	17	three_prime_utr
HEK nlambda ADAR	20_3108289_-1	UBOX5	ENSG00000185019	0.045	0.8058	0.5224	20	three_prime_utr
HEK nlambda ADAR	22_37942685_1	MICALL1	ENSG00000100139	0.0442	0.8803	0.6626	22	three_prime_utr
HEK nlambda ADAR	4_128037518_1	ABHD18	ENSG00000164074	0.0785	0.7465	0.9118	4	three_prime_utr
HEK nlambda ADAR	20_476363_-1	CSNK2A1	ENSG00000101266	0.083	0.8622	0.7413	20	three_prime_utr
HEK nlambda ADAR	7_75980358_1	POR	ENSG00000127948	0.349	0.001	0.5003	7	three_prime_utr
HEK nlambda ADAR	19_37248048_1	ZNF383	ENSG00000188283	0.0426	0.8784	0.9504	19	three_prime_utr
HEK nlambda ADAR	12_125027414_1	BRI3BP	ENSG00000184992	0.0684	0.7651	0.9408	12	three_prime_utr
HEK nlambda ADAR	19_39491240_1	TIMM50	ENSG00000105197	0.0518	0.6964	0.7427	19	three_prime_utr
HEK nlambda ADAR	1_9102204_-1	GPR157	ENSG00000180758	0.0536	0.8771	0.7975	1	three_prime_utr
HEK nlambda ADAR	14_55367576_-1	ATG14	ENSG00000126775	0.0436	0.779	0.7639	14	three_prime_utr

HEK nlambda ADAR	19_21190800_1	ZNF431	ENSG00000196705	0.0902	0.4378	0.7158	19	three_prime_utr
HEK nlambda ADAR	1_151871215_-1	THEM4	ENSG00000159445	0.0982	0.6671	0.9391	1	three_prime_utr
HEK nlambda ADAR	X_301073_1	PLCXD1	ENSG00000182378	0.05	0.9322	0.9559	X	three_prime_utr
HEK nlambda ADAR	8_38972234_1	PLEKHA2	ENSG00000169499	0.0861	0.8645	0.7595	8	three_prime_utr
HEK nlambda ADAR	1_114584376_-1	DENND2C	ENSG00000175984	0.0749	0.9901	0.9122	1	three_prime_utr
HEK nlambda ADAR	6_13790655_-1	MCUR1	ENSG00000050393	0.0452	0.8077	0.9363	6	three_prime_utr
HEK nlambda ADAR	20_3867510_1	MAVS	ENSG00000088888	0.0825	0.4846	0.9143	20	three_prime_utr
HEK nlambda ADAR	14_31094113_1	AP4S1	ENSG00000100478	0.0485	0.9344	0.9199	14	three_prime_utr
HEK nlambda ADAR	2_215944198_-1	MREG	ENSG00000118242	0.0441	0.8775	0.9335	2	three_prime_utr
HEK nlambda ADAR	X_16763375_1	SYAP1	ENSG00000169895	0.0404	1.0303	0.6971	X	three_prime_utr
HEK nlambda ADAR	3_197884399_1	LRCH3	ENSG00000186001	0.0875	0.8567	0.9355	3	three_prime_utr
HEK nlambda ADAR	17_62451250_1	METTL2A	ENSG00000087995	0.0415	0.9308	0.9294	17	three_prime_utr
HEK nlambda ADAR	7_101312953_-1	IFT22	ENSG00000128581	0.0418	0.9493	0.8306	7	three_prime_utr
HEK nlambda ADAR	16_10531455_-1	EMP2	ENSG00000213853	0.148	0.5902	0.9392	16	three_prime_utr
HEK nlambda ADAR	10_102741460_1	SFXN2	ENSG00000156398	0.1217	0.5824	0.6759	10	three_prime_utr
HEK nlambda ADAR	2_69462530_-1	AAK1	ENSG00000115977	0.0512	0.8937	0.7852	2	three_prime_utr
HEK nlambda ADAR	11_94417108_-1	MRE11	ENSG00000020922	0.0536	0.7705	0.8791	11	three_prime_utr
HEK nlambda ADAR	11_94416748_-1	MRE11	ENSG00000020922	0.2811	0.2962	0.8793	11	three_prime_utr

HEK nlambda ADAR	X_119924662_-1	NKAP	ENSG00000101882	0.0715	0.8772	0.5557	X	three_prime_utr
HEK nlambda ADAR	17_39761481_-1	IKZF3	ENSG00000161405	0.0476	0.8843	0.6877	17	three_prime_utr
HEK nlambda ADAR	1_114583767_-1	DENND2C	ENSG00000175984	0.1455	0.9542	0.8808	1	three_prime_utr
HEK nlambda ADAR	12_100626890_1	GAS2L3	ENSG00000139354	0.053	0.9524	0.942	12	three_prime_utr
HEK nlambda ADAR	7_44078429_-1	POLM	ENSG00000122678	0.0457	0.751	0.4505	7	three_prime_utr
HEK nlambda ADAR	19_57214590_1	ZNF264	ENSG00000083844	0.0477	0.9916	0.7554	19	three_prime_utr
HEK nlambda ADAR	9_78272260_1	CEP78	ENSG00000148019	0.0419	0.6709	0.7863	9	three_prime_utr
HEK nlambda ADAR	2_216100936_1	TMEM169	ENSG00000163449	0.0489	0.9583	0.6921	2	three_prime_utr
HEK nlambda ADAR	1_182379359_-1	GLUL	ENSG00000135821	0.0465	0.89	0.5484	1	three_prime_utr
HEK nlambda ADAR	4_127840099_1	HSPA4L	ENSG00000164070	0.0598	0.9005	0.9722	4	three_prime_utr
HEK nlambda ADAR	3_141947516_-1	TFDP2	ENSG00000114126	0.3341	0.1631	0.7957	3	three_prime_utr
HEK nlambda ADAR	19_44159884_1	ZNF234	ENSG00000263002	0.2442	0.7072	0.8265	19	three_prime_utr
HEK nlambda ADAR	9_128306617_-1	TRUB2	ENSG00000167112	0.0493	0.7756	0.8378	9	three_prime_utr
HEK nlambda ADAR	11_88324226_-1	CTSC	ENSG00000109861	0.0789	0.7088	0.6452	11	three_prime_utr
HEK nlambda ADAR	17_517617_-1	VPS53	ENSG00000141252	0.0542	0.4976	0.7654	17	three_prime_utr
HEK nlambda ADAR	14_64470474_1	AKAP5	ENSG00000179841	0.0793	0.8656	0.7282	14	three_prime_utr
HEK nlambda ADAR	1_15572523_-1	AGMAT	ENSG00000116771	0.0446	0.805	0.5962	1	three_prime_utr
HEK nlambda ADAR	6_149725501_-1	NUP43	ENSG00000120253	0.122	0.6489	0.9749	6	three_prime_utr

HEK nlambda ADAR	4_127836979_1	HSPA4L	ENSG00000164070	0.1394	0.7051	0.9586	4	three_prime_utr
HEK nlambda ADAR	1_202592969_-1	SYT2	ENSG00000143858	0.0595	0.9644	0.8369	1	three_prime_utr
HEK nlambda ADAR	13_45030935_1	GPALPP1	ENSG00000133114	0.3058	0.4584	0.9779	13	three_prime_utr
HEK nlambda ADAR	6_43615801_1	POLH	ENSG00000170734	0.2456	0.4639	0.8782	6	three_prime_utr
HEK nlambda ADAR	11_126349196_1	DCPS	ENSG00000110063	0.0983	0.5577	0.931	11	three_prime_utr
HEK nlambda ADAR	3_45680941_1	LIMD1	ENSG00000144791	0.2341	0.4199	0.9294	3	three_prime_utr
HEK nlambda ADAR	2_216502197_1	RPL37A	ENSG00000197756	0.0619	0.3477	0.9472	2	three_prime_utr
HEK nlambda ADAR	16_66889483_1	PDP2	ENSG00000172840	0.291	0.6463	0.9746	16	three_prime_utr
HEK nlambda ADAR	11_111728357_1	SIK2	ENSG00000170145	0.1057	0.9382	0.9374	11	three_prime_utr
HEK nlambda ADAR	14_31095495_1	AP4S1	ENSG00000100478	0.082	0.779	0.8949	14	three_prime_utr
HEK nlambda ADAR	17_39759197_-1	IKZF3	ENSG00000161405	0.1368	0.769	0.873	17	three_prime_utr
HEK nlambda ADAR	7_128505958_1	METTL2B	ENSG00000165055	0.0725	0.8614	0.9233	7	three_prime_utr
HEK nlambda ADAR	5_131201487_1	LYRM7	ENSG00000186687	0.0938	0.7818	0.9533	5	three_prime_utr
HEK nlambda ADAR	18_80159162_-1	PARD6G	ENSG00000178184	0.0893	0.4441	0.8515	18	three_prime_utr
HEK nlambda ADAR	14_64471427_1	AKAP5	ENSG00000179841	0.0538	0.9624	0.8246	14	three_prime_utr
HEK nlambda ADAR	9_78273728_1	CEP78	ENSG00000148019	0.0961	0.6409	0.9524	9	three_prime_utr
HEK nlambda ADAR	14_31094393_1	AP4S1	ENSG00000100478	0.0415	0.9029	0.7651	14	three_prime_utr
HEK nlambda ADAR	18_11883870_1	GNAL	ENSG00000141404	0.0556	1.075	0.8539	18	three_prime_utr

HEK nlambda ADAR	14_64460824_1	MTHFD1	ENSG00000100714	0.0528	0.7445	0.5548	14	three_prime_utr
HEK nlambda ADAR	13_40811400_1	SLC25A15	ENSG00000102743	0.0531	0.8162	0.9365	13	three_prime_utr
HEK nlambda ADAR	1_179074974_1	FAM20B	ENSG00000116199	0.1492	0.8241	0.991	1	three_prime_utr
HEK nlambda ADAR	8_6644097_1	MCPH1	ENSG00000147316	0.0426	0.7405	0.6495	8	three_prime_utr
HEK nlambda ADAR	12_123626049_-1	EIF2B1	ENSG00000111361	0.0601	0.9568	0.8101	12	three_prime_utr
HEK nlambda ADAR	11_8687278_1	RPL27A	ENSG00000166441	0.0513	0.5693	0.9492	11	three_prime_utr
HEK nlambda ADAR	3_142162865_-1	GK5	ENSG00000175066	0.1863	0.5849	0.9132	3	three_prime_utr
HEK nlambda ADAR	1_228180778_1	IBA57	ENSG00000181873	0.0845	0.8763	0.8485	1	three_prime_utr
Cortex 230327	17_80398189_1	RNF213	ENSG00000173821	0.0613	0.7715	0.5313	17	three_prime_utr
Cortex 230327	15_72598033_1	ARIH1	ENSG00000166233	0.1213	0.8576	0.9163	15	three_prime_utr
Cortex 230327	19_48287478_1	ZNF114	ENSG00000178150	0.1334	0.9026	0.9796	19	three_prime_utr
Cortex 230327	19_36338467_-1	ZFP14	ENSG00000142065	0.1146	0.9753	0.927	19	three_prime_utr
Cortex 230327	19_36629869_1	ZNF382	ENSG00000161298	0.0455	0.9664	0.9253	19	three_prime_utr
Cortex 230327	15_41580678_1	TYRO3	ENSG00000092445	0.0805	0.7965	0.9321	15	three_prime_utr
Cortex 230327	2_172607165_1	PDK1	ENSG00000152256	0.2609	0.5189	0.9964	2	three_prime_utr
Cortex 230327	8_38973011_1	PLEKHA2	ENSG00000169499	0.0533	0.9562	0.9291	8	three_prime_utr
Cortex 230327	14_64481201_-1	ZBTB25	ENSG00000089775	0.0694	0.956	0.8343	14	three_prime_utr
Cortex 230327	8_63060460_-1	TTPA	ENSG00000137561	0.124	0.5978	0.901	8	three_prime_utr
Cortex 230327	17_3803954_-1	NCBP3	ENSG00000074356	0.4458	0.7364	0.9906	17	three_prime_utr
Cortex 230327	13_32515025_-1	N4BP2L2	ENSG00000244754	0.2027	0.2864	0.6935	13	three_prime_utr
Cortex 230327	16_30581080_-1	ZNF785	ENSG00000197162	0.049	0.8558	0.8929	16	three_prime_utr
Cortex 230327	14_102051145_1	DYNC1H1	ENSG00000197102	0.0469	0.7161	0.4375	14	three_prime_utr
Cortex 230327	10_100554374_1	HIF1AN	ENSG00000166135	0.0448	0.9932	0.9474	10	three_prime_utr

Cortex 230327	22_23739310_-1	ZNF70	ENSG00000187792	0.0449	0.969	0.9798	22	three_prime_utr
Cortex 230327	19_2856466_1	ZNF555	ENSG00000186300	0.0408	0.9496	0.8901	19	three_prime_utr
Cortex 230327	5_75368984_-1	CERT1	ENSG00000113163	0.1033	1.0312	0.8325	5	three_prime_utr
Cortex 230327	19_52880345_-1	ZNF320	ENSG00000182986	0.0435	1.0158	0.9366	19	three_prime_utr
Cortex 230327	14_60138022_1	PCNX4	ENSG00000126773	0.0861	0.878	0.8463	14	three_prime_utr
Cortex 230327	9_94295149_1	ZNF169	ENSG00000175787	0.1432	0.7403	0.8056	9	three_prime_utr
Cortex 230327	19_57863256_1	ZNF587	ENSG00000198466	0.0525	0.997	0.9826	19	three_prime_utr
Cortex 230327	19_36632257_1	ZNF382	ENSG00000161298	0.0604	0.554	0.8778	19	three_prime_utr
Cortex 230327	10_14904103_1	SUV39H2	ENSG00000152455	0.0522	0.9843	0.9575	10	three_prime_utr
Cortex 230327	19_11065598_1	SMARCA4	ENSG00000127616	0.1964	0.6655	0.9844	19	three_prime_utr
Cortex 230327	11_67462851_-1	TMEM134	ENSG00000172663	0.0502	0.7233	0.6693	11	three_prime_utr
Cortex 230327	13_67206369_-1	PCDH9	ENSG00000184226	0.1231	0.9257	0.9875	13	three_prime_utr
Cortex 230327	19_36630815_1	ZNF382	ENSG00000161298	0.0442	0.4877	0.8513	19	three_prime_utr
Cortex 230327	11_44247456_1	EXT2	ENSG00000151348	0.0708	0.9076	0.7204	11	three_prime_utr
Cortex 230327	11_77614573_-1	CLNS1A	ENSG00000074201	0.0553	0.865	0.9514	11	three_prime_utr
Cortex 230327	1_36064124_1	AGO3	ENSG00000126070	0.0585	0.8148	0.4297	1	three_prime_utr
Cortex 230327	10_102740350_1	SFXN2	ENSG00000156398	0.1651	0.9163	0.9084	10	three_prime_utr
Cortex 230327	3_50595141_1	HEMK1	ENSG00000114735	0.1084	0.9915	0.9833	3	three_prime_utr
Cortex 230327	1_146995373_1	NBPF12	ENSG00000268043	0.064	0.5781	0.9039	1	three_prime_utr
Cortex 230327	8_141141976_1	DENND3	ENSG00000105339	0.2792	0.5482	0.9536	8	three_prime_utr
Cortex 230327	12_50148496_-1	CERS5	ENSG00000139624	0.0838	0.9096	0.8512	12	three_prime_utr
Cortex 230327	5_134967274_1	PCBD2	ENSG00000132570	0.0536	0.9529	0.9025	5	three_prime_utr
Cortex 230327	17_82112264_-1	CCDC57	ENSG00000176155	0.0647	0.7425	0.6113	17	three_prime_utr
Cortex 230327	19_58126785_-1	ZNF329	ENSG00000181894	0.0578	0.9236	0.9559	19	three_prime_utr
Cortex 230327	1_22602319_1	EPHA8	ENSG00000070886	0.0881	0.6805	0.8126	1	three_prime_utr
Cortex 230327	19_53236151_-1	ZNF677	ENSG00000197928	0.092	0.9144	0.9396	19	three_prime_utr
Cortex 230327	X_134420031_1	PHF6	ENSG00000156531	0.106	0.9083	0.9146	X	three_prime_utr

Cortex 230327	1_36069957_1	AGO3	ENSG00000126070	0.1051	0.3762	0.8179	1	three_prime_utr
Cortex 230327	17_81563344_-1	NPLOC4	ENSG00000182446	0.0655	0.8494	0.74	17	three_prime_utr
Cortex 230327	1_155182832_1	TRIM46	ENSG00000163462	0.1657	0.8058	0.9579	1	three_prime_utr
Cortex 230327	19_39427507_1	PLEKHG2	ENSG00000090924	0.1709	0.5539	0.9895	19	three_prime_utr
Cortex 230327	19_57699920_-1	ZNF154	ENSG00000179909	0.2698	0.631	0.9769	19	three_prime_utr
Cortex 230327	19_57699319_-1	ZNF154	ENSG00000179909	0.0645	0.9498	0.883	19	three_prime_utr
Cortex 230327	7_102454136_1	ORAI2	ENSG00000160991	0.1013	0.6625	0.6924	7	three_prime_utr
Cortex 230327	12_133117980_-1	ZNF891	ENSG00000214029	0.0726	0.8471	0.763	12	three_prime_utr
Cortex 230327	16_30582001_-1	ZNF785	ENSG00000197162	0.068	0.4152	0.4982	16	three_prime_utr
Cortex 230327	22_23740066_-1	ZNF70	ENSG00000187792	0.1193	0.8954	0.9442	22	three_prime_utr
Cortex 230327	20_50881729_1	BCAS4	ENSG00000124243	0.0461	0.7087	0.9551	20	three_prime_utr
Cortex 230327	10_15106085_-1	NMT2	ENSG00000152465	0.0542	0.9365	0.9732	10	three_prime_utr
Cortex 230327	19_36389332_-1	ZFP82	ENSG00000181007	0.0858	0.909	0.9467	19	three_prime_utr
Cortex 230327	19_16095862_1	TPM4	ENSG00000167460	0.0658	0.9067	0.8588	19	three_prime_utr
Cortex 230327	1_161751185_1	DUSP12	ENSG00000081721	0.0622	0.9645	0.8676	1	three_prime_utr
Cortex 230327	9_94295141_1	ZNF169	ENSG00000175787	0.0733	0.9093	0.9943	9	three_prime_utr
Cortex 230327	6_5104556_-1	LYRM4	ENSG00000214113	0.0579	0.839	0.6121	6	three_prime_utr
Cortex 230327	10_94330430_1	PLCE1	ENSG00000138193	0.1212	0.7497	0.9013	10	three_prime_utr
Cortex 230327	19_57222094_1	ZNF264	ENSG00000083844	0.1221	0.7988	0.9801	19	three_prime_utr
Cortex 230327	19_52293229_1	ZNF766	ENSG00000196214	0.1487	0.7793	0.8842	19	three_prime_utr
Cortex 230327	19_20540689_-1	ZNF737	ENSG00000237440	0.1107	0.6732	0.9475	19	three_prime_utr
Cortex 230327	2_201174616_1	CFLAR	ENSG00000003402	0.2306	0.6422	0.8476	2	three_prime_utr
Cortex 230327	3_188880273_1	LPP	ENSG00000145012	0.0554	0.9282	0.6873	3	three_prime_utr
Cortex 230327	2_203298657_1	CYP20A1	ENSG00000119004	0.0508	0.9134	0.9492	2	three_prime_utr
Cortex 230327	11_76406235_1	GVQW3	ENSG00000179240	0.1652	0.8057	0.9743	11	three_prime_utr
Cortex 230327	15_63152119_-1	RPS27L	ENSG00000185088	0.0931	0.7078	0.9007	15	three_prime_utr
Cortex 230327	18_49595574_1	LIPG	ENSG00000101670	0.0664	0.9888	0.8377	18	three_prime_utr

Cortex 230327	14_64480164_-1	ZBTB25	ENSG00000089775	0.0468	0.7762	0.4204	14	three_prime_utr
Cortex 230327	7_48103963_1	UPP1	ENSG00000183696	0.1419	0.6342	0.8796	7	three_prime_utr
Cortex 230327	6_26096169_1	HFE	ENSG00000010704	0.113	0.5981	0.5895	6	three_prime_utr
Cortex 230327	22_18092717_1	PEX26	ENSG00000215193	0.0497	0.9381	0.8919	22	three_prime_utr
Cortex 230327	14_64483349_-1	ZBTB25	ENSG00000089775	0.1029	0.8348	0.8507	14	three_prime_utr
Cortex 230327	13_51441931_-1	INTS6	ENSG00000102786	0.0429	0.907	0.5469	13	three_prime_utr
Cortex 230327	19_40032605_-1	ZNF780B	ENSG00000128000	0.1738	0.7402	0.9516	19	three_prime_utr
Cortex 230327	18_49478092_-1	C18orf32	ENSG00000177576	0.1658	0.8469	0.9298	18	three_prime_utr
Cortex 230327	12_133109968_-1	ZNF891	ENSG00000214029	0.0527	0.9957	0.9174	12	three_prime_utr
Cortex 230327	X_74584724_-1	RLIM	ENSG00000131263	0.1722	0.7269	0.8528	X	three_prime_utr
Cortex 230327	10_94332052_1	PLCE1	ENSG00000138193	0.0715	0.7275	0.5117	10	three_prime_utr
Cortex 230327	15_50558949_-1	TRPM7	ENSG00000092439	0.1193	0.9214	0.9243	15	three_prime_utr
Cortex 230327	19_57792953_1	ZNF586	ENSG00000083828	0.1263	0.8887	0.8886	19	three_prime_utr
Cortex 230327	3_150618775_1	SELENOT	ENSG00000198843	0.0521	0.9304	0.8554	3	three_prime_utr
Cortex 230327	12_133213588_1	ZNF268	ENSG00000090612	0.0402	0.7268	0.5686	12	three_prime_utr
Cortex 230327	11_8615066_-1	TRIM66	ENSG00000166436	0.0449	0.94	0.8216	11	three_prime_utr
Cortex 230327	21_43805015_1	RRP1	ENSG00000160214	0.0923	0.4921	0.4857	21	three_prime_utr
Cortex 230327	12_133113751_-1	ZNF891	ENSG00000214029	0.0627	0.9584	0.9695	12	three_prime_utr
Cortex 230327	19_57793109_1	ZNF586	ENSG00000083828	0.0471	0.8462	0.9048	19	three_prime_utr
Cortex 230327	18_47827618_-1	SMAD2	ENSG00000175387	0.1259	0.8925	0.9428	18	three_prime_utr
Cortex 230327	15_72597979_1	ARIH1	ENSG00000166233	0.1365	0.7843	0.8933	15	three_prime_utr
Cortex 230327	19_11914195_1	ZNF69	ENSG00000198429	0.2592	0.7175	0.9842	19	three_prime_utr
Cortex 230327	19_7683831_1	TRAPPC5	ENSG00000181029	0.0494	0.7713	0.557	19	three_prime_utr
Cortex 230327	12_133014684_1	ZNF26	ENSG00000198393	0.0531	0.8555	0.7853	12	three_prime_utr
Cortex 230327	5_141191029_1	PCDHB9	ENSG00000177839	0.0512	0.6538	0.5572	5	three_prime_utr
Cortex 230327	19_36633177_1	ZNF382	ENSG00000161298	0.0563	0.8169	0.8505	19	three_prime_utr
Cortex 230327	6_52493600_1	EFHC1	ENSG00000096093	0.0775	0.78	0.7966	6	three_prime_utr

Cortex 230327	14_64480198_-1	ZBTB25	ENSG00000089775	0.1104	0.8313	0.8922	14	three_prime_utr
Cortex 230327	19_37540026_1	ZNF793	ENSG00000188227	0.1032	0.8308	0.8912	19	three_prime_utr
Cortex 230327	2_32315537_1	YIPF4	ENSG00000119820	0.0955	1.0183	0.9439	2	three_prime_utr
Cortex 230327	10_80529974_1	TSPAN14	ENSG00000108219	0.0683	0.8902	0.9546	10	three_prime_utr
Cortex 230327	19_7683853_1	TRAPPC5	ENSG00000181029	0.1315	0.6785	0.9119	19	three_prime_utr
Cortex 230327	12_42234985_-1	YAF2	ENSG00000015153	0.0506	0.9549	0.9205	12	three_prime_utr
Cortex 230327	19_36633244_1	ZNF382	ENSG00000161298	0.0564	0.8973	0.9359	19	three_prime_utr
Cortex 230327	15_72599274_1	ARIH1	ENSG00000166233	0.0736	0.6366	0.7197	15	three_prime_utr
Cortex 230327	2_203446787_1	ABI2	ENSG00000138443	0.0636	0.9822	0.7801	2	three_prime_utr
Cortex 230327	15_44348797_1	CASC4	ENSG00000166734	0.0625	0.9674	0.85	15	three_prime_utr
Cortex 230327	12_133209123_1	ZNF268	ENSG00000090612	0.0866	0.9307	0.9285	12	three_prime_utr
Cortex 230327	2_32313506_1	YIPF4	ENSG00000119820	0.1873	0.8817	0.7527	2	three_prime_utr
Cortex 230327	10_14904183_1	SUV39H2	ENSG00000152455	0.0987	0.9341	0.8706	10	three_prime_utr
Cortex 230327	4_98874433_-1	EIF4E	ENSG00000151247	0.0764	0.93	0.9916	4	three_prime_utr
Cortex 230327	20_3928684_-1	RNF24	ENSG00000101236	0.0517	0.9525	0.929	20	three_prime_utr
Cortex 230327	18_49478395_-1	C18orf32	ENSG00000177576	0.0564	0.9757	0.866	18	three_prime_utr
Cortex 230327	18_50934540_1	ME2	ENSG00000082212	0.0773	0.8997	0.772	18	three_prime_utr
Cortex 230327	6_111275538_1	MFSD4B	ENSG00000173214	0.0831	0.8688	0.9425	6	three_prime_utr
Cortex 230327	9_109136781_-1	FRRS1L	ENSG00000260230	0.0598	0.5578	0.6406	9	three_prime_utr
Cortex 230327	19_53494816_1	ZNF813	ENSG00000198346	0.0556	0.8889	0.8608	19	three_prime_utr
Both	15_75936066_1	FBXO22	ENSG00000167196	0.0438	0.9152	0.8693	15	three_prime_utr
Both	22_40931590_1	XPNPEP3	ENSG00000196236	0.189	0.5272	0.9314	22	three_prime_utr
Both	19_21123539_1	ZNF714	ENSG00000160352	0.0803	0.7619	0.8806	19	three_prime_utr
Both	5_65668574_-1	SGTB	ENSG00000197860	0.0568	0.8834	0.8879	5	three_prime_utr
Both	17_40283074_1	WIPF2	ENSG00000171475	0.0401	0.9774	0.5691	17	three_prime_utr
Both	X_46499300_-1	ZNF674	ENSG00000251192	0.0404	0.7947	0.5297	X	three_prime_utr
Both	X_300734_1	PLCXD1	ENSG00000182378	0.1342	0.792	0.945	X	three_prime_utr

Both	19_12632713_1	ZNF791	ENSG00000173875	0.1021	0.8617	0.766	19	three_prime_utr
Both	13_49913804_-1	SPRYD7	ENSG00000123178	0.2271	0.3104	0.9838	13	three_prime_utr
Both	11_76384399_1	GVQW3	ENSG00000179240	0.2895	0.4792	0.6024	11	three_prime_utr
Both	17_67338650_-1	PSMD12	ENSG00000197170	0.1621	0.6071	0.9802	17	three_prime_utr
Both	2_128194449_1	UGGT1	ENSG00000136731	0.0508	0.8916	0.958	2	three_prime_utr
Both	16_10529685_-1	EMP2	ENSG00000213853	0.0945	0.908	0.9699	16	three_prime_utr
Both	3_48750128_-1	PRKAR2A	ENSG00000114302	0.0983	0.8128	0.9722	3	three_prime_utr
Both	X_302702_1	PLCXD1	ENSG00000182378	0.1142	0.9166	0.9686	X	three_prime_utr
Both	1_9584463_1	SLC25A33	ENSG00000171612	0.1346	0.5218	0.8898	1	three_prime_utr
Both	20_13819421_1	NDUFAF5	ENSG00000101247	0.0779	0.5439	0.7017	20	three_prime_utr
Both	17_37478010_1	TADA2A	ENSG00000276234	0.4911	0.2166	0.9638	17	three_prime_utr
Both	17_44800252_-1	GJC1	ENSG00000182963	0.1564	0.4757	0.8391	17	three_prime_utr
Both	1_35599887_-1	PSMB2	ENSG00000126067	0.0495	0.7998	0.9478	1	three_prime_utr
Both	16_89533427_1	SPG7	ENSG00000197912	0.0416	0.7939	0.6241	16	three_prime_utr
Both	1_19217478_-1	EMC1	ENSG00000127463	0.0535	0.7365	0.8487	1	three_prime_utr
Both	3_45680376_1	LIMD1	ENSG00000144791	0.1118	0.6346	0.9667	3	three_prime_utr
Both	7_25120867_-1	CYCS	ENSG00000172115	0.0755	0.9447	0.9349	7	three_prime_utr
Both	16_48354336_1	LONP2	ENSG00000102910	0.0882	0.6689	0.9511	16	three_prime_utr
Both	22_37942747_1	MICALL1	ENSG00000100139	0.1462	0.8298	0.8785	22	three_prime_utr
Both	12_100627486_1	GAS2L3	ENSG00000139354	0.0793	0.9055	0.9572	12	three_prime_utr
Both	14_64471358_1	AKAP5	ENSG00000179841	0.3989	0.0559	0.6843	14	three_prime_utr
Both	10_20286903_1	PLXDC2	ENSG00000120594	0.1617	0.7599	0.8135	10	three_prime_utr
Both	1_32683812_1	RBBP4	ENSG00000162521	0.0452	0.8198	0.9512	1	three_prime_utr
Both	9_112686592_-1	INIP	ENSG00000148153	0.0577	0.916	0.9636	9	three_prime_utr
Both	19_21185589_1	ZNF431	ENSG00000196705	0.0758	0.8847	0.8911	19	three_prime_utr
Both	1_1561263_-1	SSU72	ENSG00000160075	0.0661	0.8615	0.8308	1	three_prime_utr
Both	22_18100179_1	PEX26	ENSG00000215193	0.0966	0.8191	0.8557	22	three_prime_utr

Both	X_23701828_-1	ACOT9	ENSG00000123130	0.1601	0.2671	0.8477	X	three_prime_utr
Both	1_6522441_-1	NOL9	ENSG00000162408	0.0478	0.8912	0.8226	1	three_prime_utr
Both	4_165075532_-1	TMEM192	ENSG00000170088	0.1194	0.6544	0.9457	4	three_prime_utr
Both	22_18095866_1	PEX26	ENSG00000215193	0.0565	0.9953	0.7496	22	three_prime_utr
Both	21_33529631_-1	GART	ENSG00000159131	0.0575	0.6716	0.5595	21	three_prime_utr
Both	20_3929026_-1	RNF24	ENSG00000101236	0.1113	0.8515	0.7461	20	three_prime_utr
Both	19_53987167_1	CACNG8	ENSG00000142408	0.0799	0.4972	0.4769	19	three_prime_utr
Both	8_143655406_1	ZNF623	ENSG00000183309	0.1006	0.994	0.9414	8	three_prime_utr
Both	3_45680496_1	LIMD1	ENSG00000144791	0.0929	0.7812	0.9772	3	three_prime_utr
Both	1_52824013_1	ZYG11B	ENSG00000162378	0.1724	0.7792	0.9731	1	three_prime_utr
Both	20_36781369_-1	SOGA1	ENSG00000149639	0.1484	0.2068	0.4442	20	three_prime_utr
Both	16_89566542_1	RPL13	ENSG00000167526	0.0443	0.9287	0.802	16	three_prime_utr
Both	9_78273531_1	CEP78	ENSG00000148019	0.1697	0.5268	0.9273	9	three_prime_utr
Both	3_15072893_-1	RBSN	ENSG00000131381	0.3196	0.4091	0.9818	3	three_prime_utr
Both	1_28499518_1	PHACTR4	ENSG00000204138	0.1758	0.6307	0.9504	1	three_prime_utr
Both	17_39257080_-1	FBXL20	ENSG00000108306	0.1164	0.9126	0.8102	17	three_prime_utr
Both	X_123912919_1	XIAP	ENSG00000101966	0.0626	0.9631	0.9531	X	three_prime_utr
Both	15_51912376_1	TMOD3	ENSG00000138594	0.0483	0.4608	0.5385	15	three_prime_utr
Both	19_43850478_1	ZNF283	ENSG00000167637	0.1341	0.5744	0.6548	19	three_prime_utr
Both	15_82842396_-1	HOMER2	ENSG00000103942	0.2917	0.281	0.7766	15	three_prime_utr
Both	7_128506025_1	METTL2B	ENSG00000165055	0.0401	0.9295	0.6631	7	three_prime_utr
Both	19_7084479_1	ZNF557	ENSG00000130544	0.0676	0.8857	0.8826	19	three_prime_utr
Both	16_50105651_1	HEATR3	ENSG00000155393	0.0818	0.7709	0.9848	16	three_prime_utr
Both	3_40538263_1	ZNF621	ENSG00000172888	0.0748	0.9513	0.9573	3	three_prime_utr
Both	20_3929072_-1	RNF24	ENSG00000101236	0.3184	0.6069	0.9125	20	three_prime_utr
Both	16_22288088_1	EEF2K	ENSG00000103319	0.1408	0.7625	0.8805	16	three_prime_utr
Both	4_88258827_-1	PPM1K	ENSG00000163644	0.0635	0.9631	0.9542	4	three_prime_utr

Both	20_3929050_-1	RNF24	ENSG00000101236	0.1062	0.9452	0.9495	20	three_prime_utr
Both	17_2417214_-1	METTL16	ENSG00000127804	0.3387	0.1745	0.9762	17	three_prime_utr
Both	12_111895307_1	MAPKAPK5	ENSG00000089022	0.1718	0.6658	0.8695	12	three_prime_utr
Both	2_85611168_-1	C2orf68	ENSG00000168887	0.1887	0.854	0.9458	2	three_prime_utr
Both	2_206118573_-1	NDUFS1	ENSG00000023228	0.0903	0.9678	0.9154	2	three_prime_utr
Both	1_92396792_1	RPAP2	ENSG00000122484	0.0515	0.7288	0.5188	1	three_prime_utr
Both	17_81561856_-1	NPLOC4	ENSG00000182446	0.1145	0.8934	0.8313	17	three_prime_utr
Both	20_44305596_-1	FITM2	ENSG00000197296	0.1087	0.8334	0.9855	20	three_prime_utr
Both	7_44877545_-1	PURB	ENSG00000146676	0.0928	1.0437	0.8941	7	three_prime_utr
Both	1_32602180_1	ZBTB8A	ENSG00000160062	0.0699	0.8606	0.9141	1	three_prime_utr
Both	12_27423535_1	ARNTL2	ENSG00000029153	0.2704	0.526	0.9384	12	three_prime_utr
Both	17_32328700_-1	C17orf75	ENSG00000108666	0.2448	0.284	0.8544	17	three_prime_utr
Both	5_40831001_-1	RPL37	ENSG00000145592	0.1473	0.8417	0.9392	5	three_prime_utr
Both	5_126625762_1	PHAX	ENSG00000164902	0.0742	0.8954	0.9781	5	three_prime_utr
Both	7_25120223_-1	CYCS	ENSG00000172115	0.0796	0.8128	0.861	7	three_prime_utr
Both	4_17801243_-1	DCAF16	ENSG00000163257	0.0789	0.9158	0.9455	4	three_prime_utr
Both	7_25119547_-1	CYCS	ENSG00000172115	0.1554	0.7071	0.9435	7	three_prime_utr
Both	16_19702254_-1	KNOP1	ENSG00000103550	0.1586	0.7339	0.9674	16	three_prime_utr
Both	4_73057944_-1	COX18	ENSG00000163626	0.0721	0.8585	0.9619	4	three_prime_utr
Both	12_75496272_-1	KRR1	ENSG00000111615	0.0428	0.8199	0.9544	12	three_prime_utr
Both	5_40831250_-1	RPL37	ENSG00000145592	0.1509	0.8431	0.9803	5	three_prime_utr
Both	3_15072291_-1	RBSN	ENSG00000131381	0.0421	0.8646	0.9697	3	three_prime_utr
Both	12_71924455_1	TBC1D15	ENSG00000121749	0.0721	0.5409	0.6377	12	three_prime_utr
Both	14_74736296_1	FCF1	ENSG00000119616	0.0637	0.9376	0.9832	14	three_prime_utr
Both	2_113710472_-1	SLC35F5	ENSG00000115084	0.0438	0.9036	0.7125	2	three_prime_utr
Both	10_119572449_-1	TIAL1	ENSG00000151923	0.222	0.4325	0.7612	10	three_prime_utr
Both	1_9583096_1	SLC25A33	ENSG00000171612	0.1113	0.7636	0.9419	1	three_prime_utr

Both	15_34104827_1	PGBD4	ENSG00000182405	0.1402	0.9843	0.9272	15	three_prime_utr
Both	5_149981851_1	SLC26A2	ENSG00000155850	0.0475	0.9739	0.9178	5	three_prime_utr
Both	1_108932949_1	GPSM2	ENSG00000121957	0.0718	0.9546	0.9322	1	three_prime_utr
Both	11_769694_-1	GATD1	ENSG00000177225	0.0755	0.6402	0.8369	11	three_prime_utr
Both	14_55367729_-1	ATG14	ENSG00000126775	0.2485	0.5281	0.9611	14	three_prime_utr
Both	6_44450097_1	CDC5L	ENSG00000096401	0.0697	0.9214	0.9647	6	three_prime_utr
Both	16_70249335_-1	EXOSC6	ENSG00000223496	0.0821	0.7241	0.9087	16	three_prime_utr
Both	13_45030505_1	GPALPP1	ENSG00000133114	0.1255	0.7789	0.9733	13	three_prime_utr
Both	5_65668838_-1	SGTB	ENSG00000197860	0.042	0.8738	0.8461	5	three_prime_utr
Both	9_36216463_-1	GNE	ENSG00000159921	0.1015	0.3193	0.7444	9	three_prime_utr
Both	16_19702186_-1	KNOP1	ENSG00000103550	0.1252	0.8338	0.9516	16	three_prime_utr
Both	19_21291698_-1	ZNF708	ENSG00000182141	0.1279	0.5248	0.8266	19	three_prime_utr
Both	19_9561635_-1	ZNF121	ENSG00000197961	0.0655	0.9354	0.9293	19	three_prime_utr
Both	17_37515631_-1	SYNRG	ENSG00000275066	0.064	0.9353	0.9619	17	three_prime_utr
Both	15_34231317_-1	SLC12A6	ENSG00000140199	0.0806	0.8252	0.9303	15	three_prime_utr
Both	22_41600021_-1	DESI1	ENSG00000100418	0.1742	0.4536	0.9244	22	three_prime_utr
Both	4_127839569_1	HSPA4L	ENSG00000164070	0.0458	0.9153	0.8558	4	three_prime_utr
Both	1_37536077_-1	SNIP1	ENSG00000163877	0.1882	0.6879	0.9199	1	three_prime_utr
Both	12_98606011_1	SLC25A3	ENSG00000075415	0.0546	0.8449	0.9563	12	three_prime_utr
Both	17_27314428_1	WSB1	ENSG00000109046	0.0521	0.9759	0.9373	17	three_prime_utr
Both	4_127837834_1	HSPA4L	ENSG00000164070	0.0672	0.6984	0.946	4	three_prime_utr
Both	19_19680539_1	ZNF101	ENSG00000181896	0.0787	0.9508	0.8905	19	three_prime_utr
Both	20_36781318_-1	SOGA1	ENSG00000149639	0.1337	0.8936	0.9558	20	three_prime_utr
Both	7_4770284_1	FOXK1	ENSG00000164916	0.0966	1.003	0.9157	7	three_prime_utr
Both	2_206795438_1	FASTKD2	ENSG00000118246	0.0496	0.7479	0.7566	2	three_prime_utr
Both	4_165071428_-1	TMEM192	ENSG00000170088	0.1476	0.8587	0.6471	4	three_prime_utr
Both	5_157758379_1	LSM11	ENSG00000155858	0.0827	0.7409	0.9083	5	three_prime_utr

Both	2_203302743_1	CYP20A1	ENSG00000119004	0.2758	0.5214	0.8865	2	three_prime_utr
Both	12_68766687_1	SLC35E3	ENSG00000175782	0.0841	0.694	0.9254	12	three_prime_utr
Both	4_128037164_1	ABHD18	ENSG00000164074	0.0516	0.6615	0.5796	4	three_prime_utr
Both	12_116715122_-1	C12orf49	ENSG00000111412	0.0444	0.9637	0.9833	12	three_prime_utr
Both	2_203302325_1	CYP20A1	ENSG00000119004	0.0403	0.9517	0.4967	2	three_prime_utr
Both	10_86926447_1	BMPR1A	ENSG00000107779	0.092	0.9656	0.7388	10	three_prime_utr
Both	14_31094389_1	AP4S1	ENSG00000100478	0.052	0.9132	0.8459	14	three_prime_utr
Both	12_98602346_1	SLC25A3	ENSG00000075415	0.064	0.8853	0.9426	12	three_prime_utr
Both	22_18091376_1	PEX26	ENSG00000215193	0.0402	0.8988	0.8663	22	three_prime_utr
Both	11_44249814_1	EXT2	ENSG00000151348	0.0677	0.8474	0.6254	11	three_prime_utr
Both	2_113710755_-1	SLC35F5	ENSG00000115084	0.0747	0.8813	0.9059	2	three_prime_utr
Both	11_31431610_1	DNAJC24	ENSG00000170946	0.0915	0.6959	0.9747	11	three_prime_utr
Both	17_32329128_-1	C17orf75	ENSG00000108666	0.0532	0.9087	0.9084	17	three_prime_utr
Both	6_13787967_-1	MCUR1	ENSG00000050393	0.0438	0.9255	0.9738	6	three_prime_utr
Both	5_176379247_1	ARL10	ENSG00000175414	0.1202	0.6327	0.9243	5	three_prime_utr
Both	1_100189313_-1	DBT	ENSG00000137992	0.1525	0.7594	0.9326	1	three_prime_utr
Both	2_127846140_-1	POLR2D	ENSG00000144231	0.0658	0.8089	0.9045	2	three_prime_utr
Both	11_108365795_1	ATM	ENSG00000149311	0.0928	0.4539	0.9289	11	three_prime_utr
Both	19_4654443_1	TNFAIP8L1	ENSG00000185361	0.1219	0.8754	0.9093	19	three_prime_utr
Both	14_55367642_-1	ATG14	ENSG00000126775	0.1335	0.5909	0.8966	14	three_prime_utr
Both	3_57092398_-1	IL17RD	ENSG00000144730	0.2103	0.7892	0.8769	3	three_prime_utr
Both	1_220058801_-1	BPNT1	ENSG00000162813	0.0793	0.7835	0.9886	1	three_prime_utr
Both	19_39493378_1	TIMM50	ENSG00000105197	0.0451	0.8503	0.9675	19	three_prime_utr
Both	19_21122926_1	ZNF714	ENSG00000160352	0.0449	1.0004	0.6216	19	three_prime_utr
Both	4_83460435_-1	ABRAXAS1	ENSG00000163322	0.0701	0.6817	0.6934	4	three_prime_utr
Both	17_59274094_1	GDPD1	ENSG00000153982	0.2275	0.4791	0.9415	17	three_prime_utr
Both	6_7283384_-1	SSR1	ENSG00000124783	0.0611	1.0245	0.92	6	three_prime_utr

Both	11_32605050_1	EIF3M	ENSG00000149100	0.0556	1.0163	0.8066	11	three_prime_utr
Both	14_74737309_1	FCF1	ENSG00000119616	0.0743	0.8203	0.8441	14	three_prime_utr
Both	15_50703566_-1	SPPL2A	ENSG00000138600	0.1737	0.6124000000000006	0.8674	15	three_prime_utr
Both	1_3813620_-1	CEP104	ENSG00000116198	0.0842	0.9405	0.9261	1	three_prime_utr
Both	12_100627037_1	GAS2L3	ENSG00000139354	0.1634	0.7561	0.8778	12	three_prime_utr
Both	14_73699369_1	DNAL1	ENSG00000119661	0.0932	0.7996	0.9233	14	three_prime_utr
Both	20_3869787_1	MAVS	ENSG00000088888	0.059	0.7306	0.9533	20	three_prime_utr
Both	20_3873386_1	MAVS	ENSG00000088888	0.0526	0.8939	0.9648	20	three_prime_utr
Both	7_99481479_1	ZNF789	ENSG00000198556	0.092	0.9135	0.7359	7	three_prime_utr
Both	2_203300634_1	CYP20A1	ENSG00000119004	0.0601	0.9322	0.8309	2	three_prime_utr
Both	2_113710508_-1	SLC35F5	ENSG00000115084	0.1055	0.7747	0.9015	2	three_prime_utr
Both	5_132693858_-1	KIF3A	ENSG00000131437	0.125	0.6289	0.7812	5	three_prime_utr
Both	3_197886080_1	LRCH3	ENSG00000186001	0.0788	0.7482	0.7246	3	three_prime_utr
Both	18_21633111_1	SNRPD1	ENSG00000167088	0.1304	0.4222	0.7716	18	three_prime_utr
Both	11_123072375_-1	CLMP	ENSG00000166250	0.0845	0.9667	0.88	11	three_prime_utr
Both	5_179840386_-1	MRNIP	ENSG00000161010	0.0947	0.7418	0.8294	5	three_prime_utr
Both	13_52689831_1	SUGT1	ENSG00000165416	0.076	0.8353	0.8465	13	three_prime_utr
Both	17_37517002_-1	SYNRG	ENSG00000275066	0.2723	0.3677	0.9272	17	three_prime_utr
Both	17_517621_-1	VPS53	ENSG00000141252	0.054	0.909	0.8734	17	three_prime_utr
Both	4_128036574_1	ABHD18	ENSG00000164074	0.0498	0.9288	0.9266	4	three_prime_utr
Both	6_31858897_-1	NEU1	ENSG00000204386	0.04	0.7581	0.5074	6	three_prime_utr
Both	19_57214493_1	ZNF264	ENSG00000083844	0.1438	0.5562	0.5846	19	three_prime_utr
Both	2_113707631_-1	SLC35F5	ENSG00000115084	0.0952	0.7999000000000006	0.9028	2	three_prime_utr
Both	4_88260156_-1	PPM1K	ENSG00000163644	0.0679	0.9321	0.9041	4	three_prime_utr
Both	19_57216329_1	ZNF264	ENSG00000083844	0.1049	0.9084	0.9086	19	three_prime_utr
Both	7_77781864_1	RSBN1L	ENSG00000187257	0.1723	0.3586	0.8112	7	three_prime_utr
Both	X_119922975_-1	NKAP	ENSG00000101882	0.0569	0.7212	0.6173	X	three_prime_utr

Both	5_75377996_-1	CERT1	ENSG00000113163	0.0692	0.7112	0.7991	5	three_prime_utr
Both	8_94792447_1	DPY19L4	ENSG00000156162	0.1627	0.3832	0.825	8	three_prime_utr
Both	12_116711738_-1	C12orf49	ENSG00000111412	0.1386	0.6012	0.8763	12	three_prime_utr
Both	2_203298765_1	CYP20A1	ENSG00000119004	0.0456	0.6717	0.807	2	three_prime_utr
Both	22_37942642_1	MICALL1	ENSG00000100139	0.0508	0.8521	0.8739	22	three_prime_utr
Both	5_149984381_1	SLC26A2	ENSG00000155850	0.0408	0.9491	0.9943	5	three_prime_utr
Both	22_38585066_-1	FAM227A	ENSG00000184949	0.0549	0.9173	0.771	22	three_prime_utr
Both	9_36216473_-1	GNE	ENSG00000159921	0.1235	0.5063	0.8275	9	three_prime_utr
Both	3_136334872_1	PCCB	ENSG00000114054	0.0414	0.8112	0.7933	3	three_prime_utr
Both	3_57563272_1	PDE12	ENSG00000174840	0.0902	0.6248	0.7656	3	three_prime_utr
Both	3_141946784_-1	TFDP2	ENSG00000114126	0.0922	0.6569	0.9403	3	three_prime_utr
Both	22_40929291_1	XPNPEP3	ENSG00000196236	0.1122	0.822	0.7976	22	three_prime_utr
Both	5_131202042_1	LYRM7	ENSG00000186687	0.1776	0.2918	0.914	5	three_prime_utr
Both	18_21630497_1	SNRPD1	ENSG00000167088	0.1013	0.6091	0.6171	18	three_prime_utr
Both	20_3869354_1	MAVS	ENSG00000088888	0.0466	0.9124	0.8173	20	three_prime_utr
Both	3_197885385_1	LRCH3	ENSG00000186001	0.0714	0.8353	0.9126	3	three_prime_utr
Both	11_32604008_1	EIF3M	ENSG00000149100	0.1292	0.9237	0.9404	11	three_prime_utr
Both	12_93507877_1	MRPL42	ENSG00000198015	0.0459	0.8676	0.7779	12	three_prime_utr
Both	2_61050699_1	PEX13	ENSG00000162928	0.0952	0.8328	0.9613	2	three_prime_utr
Both	4_165077703_-1	TMEM192	ENSG00000170088	0.1812	0.8666	0.8905	4	three_prime_utr
Both	17_75889830_-1	TRIM65	ENSG00000141569	0.0486	1.0085	0.9689	17	three_prime_utr
Both	10_94612814_1	HELLS	ENSG00000119969	0.1001	0.5856	0.9063	10	three_prime_utr
Both	19_12671923_-1	AC010422.8	ENSG00000285589	0.1628	0.6836	0.7988	19	three_prime_utr
Both	19_7087393_1	ZNF557	ENSG00000130544	0.0462	0.9114	0.6788	19	three_prime_utr
Both	3_58323121_1	RPP14	ENSG00000163684	0.1487	0.6194	0.9695	3	three_prime_utr
Both	20_33646780_1	CBFA2T2	ENSG00000078699	0.0733	0.7769	0.6174	20	three_prime_utr
Both	2_128192922_1	UGGT1	ENSG00000136731	0.3188	0.2717	0.9404	2	three_prime_utr

Both	16_66889862_1	PDP2	ENSG00000172840	0.0904	0.7644	0.8734	16	three_prime_utr
Both	11_72288441_-1	CLPB	ENSG00000162129	0.2853	0.3036	0.9037	11	three_prime_utr
Both	18_11883942_1	GNAL	ENSG00000141404	0.1745	1.0842	0.8732	18	three_prime_utr
Both	2_202879246_-1	WDR12	ENSG00000138442	0.0511	0.9473	0.9569	2	three_prime_utr
Both	1_179860398_-1	TOR1AIP2	ENSG00000169905	0.0899	0.8332	0.6116	1	three_prime_utr
Both	5_149983828_1	SLC26A2	ENSG00000155850	0.2265	0.8224	0.9569	5	three_prime_utr
Both	18_21633123_1	SNRPD1	ENSG00000167088	0.0647	0.6843	0.8059	18	three_prime_utr
Both	16_46799752_-1	C16orf87	ENSG00000155330	0.2169	0.5315	0.7758	16	three_prime_utr
Both	11_108367086_1	ATM	ENSG00000149311	0.052	0.8485	0.9314	11	three_prime_utr
Both	9_36215242_-1	GNE	ENSG00000159921	0.1109	0.4012	0.8734	9	three_prime_utr
Both	5_131201962_1	LYRM7	ENSG00000186687	0.1636	0.4226	0.9761	5	three_prime_utr
Both	1_36070005_1	AGO3	ENSG00000126070	0.0544	0.8824	0.874	1	three_prime_utr
Both	19_21190869_1	ZNF431	ENSG00000196705	0.0853	0.9714	0.8874	19	three_prime_utr
Both	3_156541577_-1	SSR3	ENSG00000114850	0.0511	0.6066	0.9193	3	three_prime_utr
Both	3_10152932_1	VHL	ENSG00000134086	0.0559	0.9829	0.9738	3	three_prime_utr
Both	22_29506506_-1	THOC5	ENSG00000100296	0.0588	0.6559	0.5915	22	three_prime_utr
Both	12_121781395_1	TMEM120B	ENSG00000188735	0.0464	0.9267	0.9367	12	three_prime_utr
Both	18_35249541_1	ZNF397	ENSG00000186812	0.0753	0.8221	0.8689	18	three_prime_utr
Both	2_32310621_1	YIPF4	ENSG00000119820	0.0753	0.9173	0.9522	2	three_prime_utr
Both	19_9647811_-1	ZNF562	ENSG00000171466	0.0409	0.9608	0.7909	19	three_prime_utr
Both	3_101823963_1	NXPE3	ENSG00000144815	0.0563	1.007	0.8995	3	three_prime_utr
Both	5_176379261_1	ARL10	ENSG00000175414	0.1135	0.3189	0.8995	5	three_prime_utr
Both	20_33706312_-1	PXMP4	ENSG00000101417	0.0602	0.8404	0.67490000000000006	20	three_prime_utr
Both	2_113707739_-1	SLC35F5	ENSG00000115084	0.0551	0.875	0.8141	2	three_prime_utr
Both	1_3813740_-1	CEP104	ENSG00000116198	0.1126	0.6345	0.9115	1	three_prime_utr
Both	1_168249496_1	SFT2D2	ENSG00000213064	0.0421	0.9935	0.8954	1	three_prime_utr
Both	1_28498920_1	PHACTR4	ENSG00000204138	0.0959	0.8267	0.966	1	three_prime_utr

Both	14_31094758_1	AP4S1	ENSG00000100478	0.0596	0.9524	0.9218	14	three_prime_utr
Both	19_19680683_1	ZNF101	ENSG00000181896	0.2144	0.7229	0.882	19	three_prime_utr
Both	18_21632914_1	SNRPD1	ENSG00000167088	0.2794	0.4273	0.8978	18	three_prime_utr
Both	7_25121287_-1	CYCS	ENSG00000172115	0.0443	0.9546	0.9379	7	three_prime_utr
Both	12_98605177_1	SLC25A3	ENSG00000075415	0.2114	0.4196	0.9617	12	three_prime_utr
Both	12_116850906_1	RNFT2	ENSG00000135119	0.1131	0.845	0.8328	12	three_prime_utr
Both	19_23221691_-1	ZNF724	ENSG00000196081	0.0534	0.7777	0.8183	19	three_prime_utr
Both	10_94612669_1	HELLS	ENSG00000119969	0.0533	0.9129	0.9485	10	three_prime_utr
Both	19_37245037_1	ZNF383	ENSG00000188283	0.078	0.7272	0.6526	19	three_prime_utr
Both	12_131945730_1	PUS1	ENSG00000177192	0.0577	0.8771	0.7531	12	three_prime_utr
Both	2_32308548_1	YIPF4	ENSG00000119820	0.0699	0.6805	0.8418	2	three_prime_utr
Both	1_225491278_-1	ENAH	ENSG00000154380	0.0793	0.7005	0.6864	1	three_prime_utr
Both	6_111271882_1	MFSD4B	ENSG00000173214	0.0517	0.9173	0.8483	6	three_prime_utr
Both	15_64140639_1	SNX1	ENSG00000028528	0.1313	0.8561	0.9536	15	three_prime_utr
Both	2_203303754_1	CYP20A1	ENSG00000119004	0.0774	0.6478	0.7087	2	three_prime_utr

Table 5 Differential expression analysis (DESeq2) of Monocyes stimulated with LPS

For brevity the data has been filtered to sites with padj<0.05 and -2>log2foldchange>2.

Contrast	Gene ID	baseMean	log2FoldChange	lfcSE	shrunkLFC	pvalue	padj	Gene Name
2H-0H	ENSG00000002587	12609.618	-5.662	0.088	-5.662	0.000	0.000	HS3ST1
2H-0H	ENSG00000170345	4379.518	-4.594	0.079	-4.594	0.000	0.000	FOS
2H-0H	ENSG00000140030	3289.987	-4.266	0.090	-4.266	0.000	0.000	GPR65
2H-0H	ENSG00000117036	11172.537	4.037	0.093	4.037	0.000	0.000	ETV3
2H-0H	ENSG00000023445	32533.844	4.044	0.095	4.044	0.000	0.000	BIRC3
2H-0H	ENSG00000205189	1110.172	4.102	0.107	4.102	0.000	0.000	ZBTB10
2H-0H	ENSG00000125347	15547.597	4.125	0.068	4.125	0.000	0.000	IRF1
2H-0H	ENSG00000125733	4296.785	4.132	0.074	4.132	0.000	0.000	TRIP10
2H-0H	ENSG00000136514	1967.192	4.145	0.088	4.145	0.000	0.000	RTP4
2H-0H	ENSG00000145779	56238.578	4.148	0.062	4.148	0.000	0.000	TNFAIP8
2H-0H	ENSG00000121858	10453.628	4.194	0.090	4.194	0.000	0.000	TNFSF10
2H-0H	ENSG00000184014	2337.207	4.321	0.107	4.321	0.000	0.000	DENND5A
2H-0H	ENSG00000139725	2956.076	4.343	0.081	4.343	0.000	0.000	RHOF
2H-0H	ENSG00000185215	14293.328	4.365	0.082	4.365	0.000	0.000	TNFAIP2
2H-0H	ENSG00000183484	6535.771	4.383	0.056	4.383	0.000	0.000	GPR132
2H-0H	ENSG00000164136	2620.759	4.389	0.100	4.389	0.000	0.000	IL15
2H-0H	ENSG00000117228	47436.532	4.422	0.076	4.422	0.000	0.000	GBP1
2H-0H	ENSG00000151726	49643.687	4.436	0.098	4.436	0.000	0.000	ACSL1
2H-0H	ENSG00000184588	6720.511	4.456	0.059	4.456	0.000	0.000	PDE4B
2H-0H	ENSG00000285441	16399.973	4.496	0.108	4.496	0.000	0.000	SOD2
2H-0H	ENSG00000135604	11094.992	4.553	0.098	4.553	0.000	0.000	STX11
2H-0H	ENSG00000123689	55459.349	4.558	0.100	4.558	0.000	0.000	G0S2
2H-0H	ENSG00000187608	187462.202	4.574	0.092	4.574	0.000	0.000	ISG15
2H-0H	ENSG00000104432	1139.914	4.643	0.124	4.643	0.000	0.000	IL7
2H-0H	ENSG00000131203	64101.670	4.753	0.068	4.753	0.000	0.000	IDO1
2H-0H	ENSG00000138646	18520.874	4.753	0.099	4.753	0.000	0.000	HERC5
2H-0H	ENSG00000162645	17760.097	4.782	0.070	4.782	0.000	0.000	GBP2
2H-0H	ENSG00000104312	13302.721	4.793	0.063	4.793	0.000	0.000	RIPK2
2H-0H	ENSG00000026751	95133.604	4.814	0.101	4.814	0.000	0.000	SLAMF7
2H-0H	ENSG00000144802	26456.675	4.848	0.088	4.848	0.000	0.000	NFKBIZ
2H-0H	ENSG00000163874	14271.140	4.864	0.097	4.864	0.000	0.000	ZC3H12A
2H-0H	ENSG00000107104	3193.176	4.918	0.119	4.918	0.000	0.000	KANK1
2H-0H	ENSG00000099860	14927.309	4.944	0.063	4.944	0.000	0.000	GADD45B
2H-0H	ENSG00000134470	7645.099	5.013	0.092	5.013	0.000	0.000	IL15RA
2H-0H	ENSG00000159200	3420.944	5.020	0.085	5.020	0.000	0.000	RCAN1

2H-0H	ENSG00000170542	38582.781	5.039	0.101	5.039	0.000	0.000	SERPINB9
2H-0H	ENSG00000153898	3133.356	5.045	0.117	5.045	0.000	0.000	MCOLN2
2H-0H	ENSG00000197329	21369.894	5.060	0.066	5.060	0.000	0.000	PELI1
2H-0H	ENSG00000172183	34650.706	5.115	0.078	5.115	0.000	0.000	ISG20
2H-0H	ENSG00000105939	30245.854	5.236	0.075	5.236	0.000	0.000	ZC3HAV1
2H-0H	ENSG00000168389	5675.495	5.268	0.097	5.268	0.000	0.000	MFSD2A
2H-0H	ENSG00000049249	4207.115	5.318	0.135	5.318	0.000	0.000	TNFRSF9
2H-0H	ENSG00000232810	91317.520	5.319	0.054	5.319	0.000	0.000	TNF
2H-0H	ENSG00000089351	5267.991	5.346	0.124	5.346	0.000	0.000	GRAMD1A
2H-0H	ENSG00000118503	47384.517	5.455	0.130	5.455	0.000	0.000	TNFAIP3
2H-0H	ENSG00000081041	68969.873	5.461	0.107	5.461	0.000	0.000	CXCL2
2H-0H	ENSG00000078081	7768.475	5.477	0.109	5.477	0.000	0.000	LAMP3
2H-0H	ENSG00000135114	44581.118	5.486	0.055	5.486	0.000	0.000	OASL
2H-0H	ENSG00000119917	130253.833	5.516	0.069	5.516	0.000	0.000	IFIT3
2H-0H	ENSG00000109756	9118.414	5.623	0.127	5.623	0.000	0.000	RAPGEF2
2H-0H	ENSG00000162654	16541.234	5.650	0.085	5.650	0.000	0.000	GBP4
2H-0H	ENSG00000123700	8458.045	5.695	0.083	5.695	0.000	0.000	KCNJ2
2H-0H	ENSG00000120217	11894.744	5.697	0.117	5.697	0.000	0.000	CD274
2H-0H	ENSG00000128271	6503.071	5.710	0.099	5.710	0.000	0.000	ADORA2A
2H-0H	ENSG00000140379	28405.208	5.716	0.082	5.716	0.000	0.000	BCL2A1
2H-0H	ENSG00000149289	4184.537	5.726	0.106	5.726	0.000	0.000	ZC3H12C
2H-0H	ENSG00000137393	10371.174	5.981	0.099	5.981	0.000	0.000	RNF144B
2H-0H	ENSG00000243649	4967.446	5.996	0.160	5.996	0.000	0.000	CFB
2H-0H	ENSG00000131979	1421.850	6.067	0.155	6.067	0.000	0.000	GCH1
2H-0H	ENSG00000253320	4155.495	6.134	0.144	6.134	0.000	0.000	AZIN1-AS1
2H-0H	ENSG00000234883	24285.234	6.255	0.097	6.255	0.000	0.000	MIR155HG
2H-0H	ENSG00000163739	247998.451	6.423	0.162	6.423	0.000	0.000	CXCL1
2H-0H	ENSG00000163734	194750.698	6.517	0.140	6.517	0.000	0.000	CXCL3
2H-0H	ENSG00000169245	119889.974	6.580	0.117	6.580	0.000	0.000	CXCL10
2H-0H	ENSG00000119922	171334.813	6.674	0.075	6.674	0.000	0.000	IFIT2
2H-0H	ENSG00000271503	137138.905	6.724	0.063	6.724	0.000	0.000	CCL5
2H-0H	ENSG00000141682	31776.105	6.831	0.066	6.831	0.000	0.000	PMAIP1
2H-0H	ENSG00000125538	406650.854	6.881	0.100	6.881	0.000	0.000	IL1B
2H-0H	ENSG00000166670	1616.387	6.958	0.177	6.958	0.000	0.000	MMP10
2H-0H	ENSG00000197272	6763.347	7.143	0.147	7.143	0.000	0.000	IL27
2H-0H	ENSG00000115008	4893.793	7.145	0.132	7.145	0.000	0.000	IL1A
2H-0H	ENSG00000169248	7858.924	7.643	0.159	7.643	0.000	0.000	CXCL11
2H-0H	ENSG00000117525	8301.128	7.677	0.097	7.677	0.000	0.000	F3
2H-0H	ENSG00000277632	681845.152	7.806	0.099	7.806	0.000	0.000	CCL3

2H-0H	ENSG00000123610	64667.055	8.178	0.104	8.178	0.000	0.000	TNFAIP6
2H-0H	ENSG00000073756	21640.883	8.395	0.150	8.395	0.000	0.000	PTGS2
2H-0H	ENSG00000255398	4595.526	8.774	0.179	8.774	0.000	0.000	HCAR3
2H-0H	ENSG00000276085	243846.112	8.847	0.089	8.847	0.000	0.000	CCL3L1
2H-0H	ENSG00000102794	23702.223	9.039	0.112	9.039	0.000	0.000	ACOD1
2H-0H	ENSG00000275302	461015.420	9.121	0.091	9.121	0.000	0.000	CCL4
2H-0H	ENSG00000136244	17748.154	9.212	0.142	9.212	0.000	0.000	IL6
2H-0H	ENSG00000276070	353729.863	9.621	0.105	9.621	0.000	0.000	CCL4L2
2H-0H	ENSG00000115009	61295.772	9.967	0.150	9.967	0.000	0.000	CCL20
2H-0H	ENSG00000205755	1248.343	4.496	0.166	4.496	2.84E-163	1.03E-161	CRLF2
2H-0H	ENSG00000092067	117.509	-4.434	0.343	-4.434	1.10E-39	1.04E-38	CEBPE
2H-0H	ENSG00000166016	569.228	5.185	0.177	5.185	2.44E-190	1.06E-188	ABTB2
2H-0H	ENSG00000271614	1727.734	4.047	0.173	4.047	4.11E-123	1.14E-121	ATP2B1-AS1
2H-0H	ENSG00000134070	7318.660	4.991	0.165	4.991	2.36E-203	1.16E-201	IRAK2
2H-0H	ENSG00000172548	698.034	7.494	0.230	7.494	2.20E-235	1.28E-233	NIPAL4
2H-0H	ENSG00000169242	45.320	5.383	0.512	5.383	1.91E-27	1.30E-26	EFNA1
2H-0H	ENSG00000171811	78.049	4.490	0.364	4.490	1.52E-36	1.32E-35	CFAP46
2H-0H	ENSG00000154451	14028.408	6.663	0.215	6.663	2.60E-213	1.33E-211	GBP5
2H-0H	ENSG00000253522	844.746	4.458	0.152	4.458	3.15E-190	1.36E-188	MIR3142HG
2H-0H	ENSG00000282608	1265.186	-7.050	0.250	-7.050	3.51E-176	1.38E-174	ADORA3
2H-0H	ENSG00000105711	245.741	4.263	0.270	4.263	1.11E-57	1.48E-56	SCN1B
2H-0H	ENSG00000172602	503.240	7.177	0.251	7.177	3.93E-181	1.60E-179	RND1
2H-0H	ENSG00000166923	1209.525	4.038	0.134	4.038	3.41E-200	1.61E-198	GREM1
2H-0H	ENSG00000140511	816.261	4.660	0.195	4.660	5.72E-128	1.63E-126	HAPLN3
2H-0H	ENSG00000134460	2344.532	4.352	0.126	4.352	2.45E-261	1.64E-259	IL2RA
2H-0H	ENSG00000136826	615.415	4.334	0.130	4.334	2.86E-246	1.77E-244	KLF4
2H-0H	ENSG00000122641	12921.064	6.445	0.370	6.445	1.12E-69	1.78E-68	INHBA
2H-0H	ENSG00000171855	1214.093	8.353	0.333	8.353	5.61E-141	1.79E-139	IFNB1
2H-0H	ENSG00000205856	266.648	4.044	0.205	4.044	9.34E-88	1.83E-86	C22orf42
2H-0H	ENSG00000105246	4390.908	5.086	0.136	5.086	2.17E-305	1.83E-303	EBI3
2H-0H	ENSG00000120337	219.995	5.156	0.228	5.156	7.21E-115	1.83E-113	TNFSF18
2H-0H	ENSG00000250274	274.190	4.570	0.198	4.570	7.24E-120	1.94E-118	AC034199.1
2H-0H	ENSG00000050730	26471.684	5.726	0.288	5.726	9.73E-90	1.97E-88	TNIP3
2H-0H	ENSG00000253123	467.247	4.109	0.171	4.109	6.99E-129	2.00E-127	AC091182.1
2H-0H	ENSG00000186827	5014.738	4.109	0.163	4.109	6.25E-142	2.02E-140	TNFRSF4
2H-0H	ENSG00000125430	9668.521	5.919	0.221	5.919	5.88E-159	2.09E-157	HS3ST3B1
2H-0H	ENSG00000119139	634.425	4.158	0.138	4.158	4.56E-201	2.17E-199	TJP2
2H-0H	ENSG00000143333	1716.927	4.688	0.134	4.688	3.08E-269	2.18E-267	RGS16
2H-0H	ENSG00000253831	889.036	4.110	0.178	4.110	8.58E-120	2.29E-118	ETV3L

2H-0H	ENSG00000184557	10409.754	6.033	0.274	6.033	9.60E-109	2.30E-107	SOCS3
2H-0H	ENSG00000274736	1033.644	4.635	0.156	4.635	5.15E-196	2.34E-194	CCL23
2H-0H	ENSG00000169429	957877.722	5.669	0.157	5.669	3.29E-286	2.59E-284	CXCL8
2H-0H	ENSG00000245848	295.355	-4.181	0.197	-4.181	1.21E-101	2.71E-100	CEBPA
2H-0H	ENSG00000121594	3131.334	5.059	0.147	5.059	4.03E-262	2.72E-260	CD80
2H-0H	ENSG00000163661	2000.863	4.501	0.152	4.501	6.12E-195	2.75E-193	PTX3
2H-0H	ENSG00000120436	1755.796	6.420	0.206	6.420	5.44E-214	2.79E-212	GPR31
2H-0H	ENSG00000110944	773.865	4.919	0.164	4.919	5.97E-200	2.80E-198	IL23A
2H-0H	ENSG00000182782	3118.059	9.377	0.268	9.377	4.04E-270	2.88E-268	HCAR2
2H-0H	ENSG00000155324	301.999	4.022	0.159	4.022	8.93E-143	2.89E-141	GRAMD2B
2H-0H	ENSG00000266709	143.733	5.239	0.373	5.239	2.69E-46	2.94E-45	AC005224.3
2H-0H	ENSG00000253227	389.675	5.457	0.191	5.457	7.57E-182	3.10E-180	AC090192.2
2H-0H	ENSG00000131459	175.645	4.220	0.279	4.220	2.53E-53	3.17E-52	GFPT2
2H-0H	ENSG00000166165	3861.304	5.014	0.142	5.014	4.74E-273	3.46E-271	CKB
2H-0H	ENSG00000251136	1018.147	4.186	0.181	4.186	1.45E-119	3.88E-118	AF117829.1
2H-0H	ENSG00000251230	15100.069	6.723	0.227	6.723	8.68E-195	3.88E-193	MIR3945HG
2H-0H	ENSG00000165474	3489.483	5.260	0.249	5.260	1.84E-100	4.08E-99	GJB2
2H-0H	ENSG00000134321	93857.564	4.163	0.120	4.163	6.32E-266	4.33E-264	RSAD2
2H-0H	ENSG00000158050	4502.568	4.444	0.418	4.444	6.37E-28	4.36E-27	DUSP2
2H-0H	ENSG00000164683	560.976	4.470	0.175	4.470	1.35E-144	4.45E-143	HEY1
2H-0H	ENSG00000137757	982.758	4.398	0.162	4.398	1.28E-163	4.67E-162	CASP5
2H-0H	ENSG00000145777	309.564	7.761	0.352	7.761	1.94E-109	4.67E-108	TSLP
2H-0H	ENSG00000101384	366.443	4.395	0.157	4.395	1.22E-173	4.74E-172	JAG1
2H-0H	ENSG00000105855	11817.836	4.157	0.131	4.157	1.04E-221	5.67E-220	ITGB8
2H-0H	ENSG00000255874	733.093	6.789	0.215	6.789	1.10E-220	5.95E-219	LINC00346
2H-0H	ENSG00000138378	5160.835	4.157	0.139	4.157	1.33E-198	6.09E-197	STAT4
2H-0H	ENSG00000188389	227.288	4.323	0.185	4.323	2.24E-122	6.15E-121	PDCD1
2H-0H	ENSG00000104213	1888.722	5.476	0.182	5.476	1.33E-200	6.31E-199	PDGFRL
2H-0H	ENSG00000242048	1944.856	4.125	0.260	4.125	5.06E-58	6.82E-57	AC093583.1
2H-0H	ENSG00000078401	387.575	4.370	0.197	4.370	2.89E-110	7.02E-109	EDN1
2H-0H	ENSG00000167695	168.193	4.334	0.189	4.334	2.77E-117	7.19E-116	TLCD3A
2H-0H	ENSG00000150687	268.324	4.248	0.223	4.248	3.96E-82	7.29E-81	PRSS23
2H-0H	ENSG00000155011	211.328	5.134	0.247	5.134	3.44E-97	7.34E-96	DKK2
2H-0H	ENSG00000088320	44.604	5.925	0.519	5.925	9.65E-32	7.45E-31	REM1
2H-0H	ENSG00000162433	4541.475	4.515	0.273	4.515	6.29E-63	9.18E-62	AK4
2H-0H	ENSG00000041982	410.129	4.709	0.271	4.709	6.07E-69	9.56E-68	TNC
2H-0H	ENSG00000198719	220.149	4.536	0.261	4.536	6.14E-69	9.65E-68	DLL1
2H-0H	ENSG00000108700	99365.331	4.552	0.143	4.552	1.78E-223	9.81E-222	CCL8
6H-0H	ENSG00000143110	10408.043	-4.520	0.107	-4.520	0.000	0.000	C1orf162

6H-0H	ENSG00000117525	8301.128	4.030	0.099	4.030	0.000	0.000	F3
6H-0H	ENSG00000137959	9315.540	4.064	0.105	4.064	0.000	0.000	IFI44L
6H-0H	ENSG00000135842	8274.161	4.064	0.101	4.064	0.000	0.000	NIBAN1
6H-0H	ENSG00000137628	5760.956	4.083	0.103	4.083	0.000	0.000	DDX60
6H-0H	ENSG00000134602	2635.355	4.104	0.069	4.104	0.000	0.000	STK26
6H-0H	ENSG00000056972	1183.501	4.109	0.099	4.109	0.000	0.000	TRAF3IP2
6H-0H	ENSG00000101916	1120.919	4.112	0.101	4.112	0.000	0.000	TLR8
6H-0H	ENSG00000136689	96893.244	4.121	0.061	4.121	0.000	0.000	IL1RN
6H-0H	ENSG00000168389	5675.495	4.132	0.098	4.132	0.000	0.000	MFSD2A
6H-0H	ENSG00000172594	1143.921	4.212	0.091	4.212	0.000	0.000	SMPDL3A
6H-0H	ENSG00000171522	5563.745	4.287	0.107	4.287	0.000	0.000	PTGER4
6H-0H	ENSG00000170525	14012.123	4.316	0.104	4.316	0.000	0.000	PFKFB3
6H-0H	ENSG00000119917	130253.833	4.334	0.069	4.334	0.000	0.000	IFIT3
6H-0H	ENSG00000136514	1967.192	4.385	0.088	4.385	0.000	0.000	RTP4
6H-0H	ENSG00000140379	28405.208	4.404	0.083	4.404	0.000	0.000	BCL2A1
6H-0H	ENSG00000101460	2670.733	4.470	0.095	4.470	0.000	0.000	MAP1LC3A
6H-0H	ENSG00000099860	14927.309	4.479	0.063	4.479	0.000	0.000	GADD45B
6H-0H	ENSG00000184588	6720.511	4.517	0.059	4.517	0.000	0.000	PDE4B
6H-0H	ENSG00000197329	21369.894	4.564	0.066	4.564	0.000	0.000	PELI1
6H-0H	ENSG00000010818	2024.026	4.605	0.112	4.605	0.000	0.000	HIVEP2
6H-0H	ENSG00000138642	5878.802	4.645	0.068	4.645	0.000	0.000	HERC6
6H-0H	ENSG00000165949	1577.870	4.696	0.103	4.696	0.000	0.000	IFI27
6H-0H	ENSG00000162645	17760.097	4.705	0.070	4.705	0.000	0.000	GBP2
6H-0H	ENSG00000108771	4894.264	4.733	0.071	4.733	0.000	0.000	DHX58
6H-0H	ENSG00000135114	44581.118	4.807	0.055	4.807	0.000	0.000	OASL
6H-0H	ENSG00000101187	1420.503	4.808	0.112	4.808	0.000	0.000	SLCO4A1
6H-0H	ENSG00000128284	3275.575	4.867	0.064	4.867	0.000	0.000	APOL3
6H-0H	ENSG00000123689	55459.349	4.924	0.100	4.924	0.000	0.000	G0S2
6H-0H	ENSG00000121858	10453.628	4.947	0.089	4.947	0.000	0.000	TNFSF10
6H-0H	ENSG00000086300	13446.577	4.948	0.097	4.948	0.000	0.000	SNX10
6H-0H	ENSG00000106701	4950.809	4.949	0.095	4.949	0.000	0.000	FSD1L
6H-0H	ENSG00000117228	47436.532	4.997	0.076	4.997	0.000	0.000	GBP1
6H-0H	ENSG00000151726	49643.687	5.033	0.098	5.033	0.000	0.000	ACSL1
6H-0H	ENSG00000128383	22244.347	5.065	0.119	5.065	0.000	0.000	APOBEC3A
6H-0H	ENSG00000164136	2620.759	5.115	0.099	5.115	0.000	0.000	IL15
6H-0H	ENSG00000138646	18520.874	5.125	0.099	5.125	0.000	0.000	HERC5
6H-0H	ENSG00000187608	187462.202	5.145	0.092	5.145	0.000	0.000	ISG15
6H-0H	ENSG00000285441	16399.973	5.146	0.108	5.146	0.000	0.000	SOD2
6H-0H	ENSG00000169245	119889.974	5.147	0.117	5.147	0.000	0.000	CXCL10

6H-0H	ENSG00000134326	5082.369	5.171	0.120	5.171	0.000	0.000	CMPK2
6H-0H	ENSG00000185885	4903.850	5.193	0.128	5.193	0.000	0.000	IFITM1
6H-0H	ENSG00000184979	9987.071	5.237	0.090	5.237	0.000	0.000	USP18
6H-0H	ENSG00000104432	1139.914	5.337	0.123	5.337	0.000	0.000	IL7
6H-0H	ENSG00000049249	4207.115	5.341	0.135	5.341	0.000	0.000	TNFRSF9
6H-0H	ENSG00000145287	2413.342	5.368	0.102	5.368	0.000	0.000	PLAC8
6H-0H	ENSG00000128271	6503.071	5.405	0.099	5.405	0.000	0.000	ADORA2A
6H-0H	ENSG00000170542	38582.781	5.441	0.101	5.441	0.000	0.000	SERPINB9
6H-0H	ENSG00000102524	29275.856	5.465	0.075	5.465	0.000	0.000	TNFSF13B
6H-0H	ENSG00000137393	10371.174	5.560	0.099	5.560	0.000	0.000	RNF144B
6H-0H	ENSG00000138378	5160.835	5.598	0.138	5.598	0.000	0.000	STAT4
6H-0H	ENSG00000174837	789.252	5.649	0.128	5.649	0.000	0.000	ADGRE1
6H-0H	ENSG00000168685	28002.502	5.651	0.106	5.651	0.000	0.000	IL7R
6H-0H	ENSG00000105855	11817.836	5.758	0.131	5.758	0.000	0.000	ITGB8
6H-0H	ENSG00000171049	1916.493	5.759	0.131	5.759	0.000	0.000	FPR2
6H-0H	ENSG00000234883	24285.234	5.768	0.097	5.768	0.000	0.000	MIR155HG
6H-0H	ENSG00000136960	779.780	5.780	0.136	5.780	0.000	0.000	ENPP2
6H-0H	ENSG00000182541	1590.193	5.804	0.099	5.804	0.000	0.000	LIMK2
6H-0H	ENSG00000117090	1752.672	5.845	0.118	5.845	0.000	0.000	SLAMF1
6H-0H	ENSG00000115008	4893.793	5.990	0.133	5.990	0.000	0.000	IL1A
6H-0H	ENSG00000125538	406650.854	6.076	0.100	6.076	0.000	0.000	IL1B
6H-0H	ENSG00000073756	21640.883	6.152	0.150	6.152	0.000	0.000	PTGS2
6H-0H	ENSG00000134321	93857.564	6.192	0.120	6.192	0.000	0.000	RSAD2
6H-0H	ENSG00000163739	247998.451	6.227	0.162	6.227	0.000	0.000	CXCL1
6H-0H	ENSG00000131979	1421.850	6.318	0.154	6.318	0.000	0.000	GCH1
6H-0H	ENSG00000089351	5267.991	6.332	0.123	6.332	0.000	0.000	GRAMD1A
6H-0H	ENSG00000186827	5014.738	6.351	0.162	6.351	0.000	0.000	TNFRSF4
6H-0H	ENSG00000134470	7645.099	6.368	0.092	6.368	0.000	0.000	IL15RA
6H-0H	ENSG00000026751	95133.604	6.375	0.101	6.375	0.000	0.000	SLAMF7
6H-0H	ENSG00000010030	1392.680	6.376	0.134	6.376	0.000	0.000	ETV7
6H-0H	ENSG00000137757	982.758	6.478	0.159	6.478	0.000	0.000	CASP5
6H-0H	ENSG00000162654	16541.234	6.587	0.085	6.587	0.000	0.000	GBP4
6H-0H	ENSG00000166670	1616.387	6.631	0.177	6.631	0.000	0.000	MMP10
6H-0H	ENSG00000253831	889.036	6.642	0.173	6.642	0.000	0.000	ETV3L
6H-0H	ENSG00000205755	1248.343	6.654	0.163	6.654	0.000	0.000	CRLF2
6H-0H	ENSG00000123700	8458.045	6.671	0.083	6.671	0.000	0.000	KCNJ2
6H-0H	ENSG00000172183	34650.706	6.674	0.078	6.674	0.000	0.000	ISG20
6H-0H	ENSG00000277632	681845.152	6.858	0.099	6.858	0.000	0.000	CCL3
6H-0H	ENSG00000120217	11894.744	6.946	0.117	6.946	0.000	0.000	CD274

6H-0H	ENSG00000004468	1997.236	6.950	0.126	6.950	0.000	0.000	CD38
6H-0H	ENSG00000105246	4390.908	7.027	0.136	7.027	0.000	0.000	EBI3
6H-0H	ENSG00000121594	3131.334	7.099	0.146	7.099	0.000	0.000	CD80
6H-0H	ENSG00000078081	7768.475	7.144	0.109	7.144	0.000	0.000	LAMP3
6H-0H	ENSG00000276085	243846.112	7.154	0.089	7.154	0.000	0.000	CCL3L1
6H-0H	ENSG00000134460	2344.532	7.160	0.124	7.160	0.000	0.000	IL2RA
6H-0H	ENSG00000255398	4595.526	7.174	0.180	7.174	0.000	0.000	HCAR3
6H-0H	ENSG00000153898	3133.356	7.307	0.116	7.307	0.000	0.000	MCOLN2
6H-0H	ENSG00000271503	137138.905	7.592	0.063	7.592	0.000	0.000	CCL5
6H-0H	ENSG00000108700	99365.331	7.673	0.143	7.673	0.000	0.000	CCL8
6H-0H	ENSG00000166165	3861.304	7.701	0.141	7.701	0.000	0.000	CKB
6H-0H	ENSG00000104213	1888.722	7.824	0.180	7.824	0.000	0.000	PDGFRL
6H-0H	ENSG00000115009	61295.772	7.837	0.150	7.837	0.000	0.000	CCL20
6H-0H	ENSG00000275302	461015.420	7.894	0.091	7.894	0.000	0.000	CCL4
6H-0H	ENSG00000276070	353729.863	7.941	0.105	7.941	0.000	0.000	CCL4L2
6H-0H	ENSG00000169248	7858.924	7.961	0.159	7.961	0.000	0.000	CXCL11
6H-0H	ENSG00000197272	6763.347	8.059	0.147	8.059	0.000	0.000	IL27
6H-0H	ENSG00000120436	1755.796	8.404	0.205	8.404	0.000	0.000	GPR31
6H-0H	ENSG00000243649	4967.446	8.584	0.159	8.584	0.000	0.000	CFB
6H-0H	ENSG00000251230	15100.069	8.647	0.226	8.647	0.000	0.000	MIR3945HG
6H-0H	ENSG00000123610	64667.055	8.811	0.104	8.811	0.000	0.000	TNFAIP6
6H-0H	ENSG00000136244	17748.154	8.879	0.142	8.879	0.000	0.000	IL6
6H-0H	ENSG00000131203	64101.670	9.280	0.067	9.280	0.000	0.000	IDO1
6H-0H	ENSG00000102794	23702.223	9.693	0.112	9.693	0.000	0.000	ACOD1
6H-0H	ENSG00000166523	10612.295	4.054	0.153	4.054	4.35E-156	1.00E-154	CLEC4E
6H-0H	ENSG00000150938	1524.969	4.176	0.155	4.176	4.27E-162	1.03E-160	CRIM1
6H-0H	ENSG00000205856	266.648	4.582	0.204	4.582	6.27E-114	1.05E-112	C22orf42
6H-0H	ENSG00000197093	942.311	-4.997	0.180	-4.997	4.33E-172	1.10E-170	GAL3ST4
6H-0H	ENSG00000184557	10409.754	6.180	0.275	6.180	6.70E-114	1.12E-112	SOCS3
6H-0H	ENSG00000092068	8254.861	-6.037	0.231	-6.037	5.32E-153	1.20E-151	SLC7A8
6H-0H	ENSG00000112303	600.318	4.634	0.129	4.634	2.49E-283	1.21E-281	VNN2
6H-0H	ENSG00000124875	3228.416	5.074	0.148	5.074	2.86E-260	1.22E-258	CXCL6
6H-0H	ENSG00000282608	1265.186	-7.101	0.250	-7.101	4.71E-179	1.28E-177	ADORA3
6H-0H	ENSG00000124216	219.320	4.646	0.213	4.646	8.12E-108	1.28E-106	SNAI1
6H-0H	ENSG00000198719	220.149	5.155	0.260	5.155	9.67E-90	1.31E-88	DLL1
6H-0H	ENSG00000165474	3489.483	8.272	0.248	8.272	3.39E-246	1.34E-244	GJB2
6H-0H	ENSG00000250510	271.614	-4.239	0.188	-4.239	8.21E-115	1.38E-113	GPR162
6H-0H	ENSG00000182782	3118.059	8.148	0.268	8.148	4.66E-204	1.45E-202	HCAR2
6H-0H	ENSG00000156234	86.881	6.416	0.380	6.416	1.42E-65	1.46E-64	CXCL13

6H-0H	ENSG00000197646	416.516	4.423	0.162	4.423	6.35E-166	1.56E-164	PDCD1LG2
6H-0H	ENSG00000163661	2000.863	5.534	0.151	5.534	3.12E-296	1.59E-294	PTX3
6H-0H	ENSG00000184489	112.024	4.598	0.263	4.598	1.45E-70	1.60E-69	PTP4A3
6H-0H	ENSG00000188372	329.281	4.269	0.162	4.269	7.09E-155	1.61E-153	ZP3
6H-0H	ENSG00000134072	3611.669	-4.120	0.294	-4.120	2.12E-46	1.63E-45	CAMK1
6H-0H	ENSG00000180767	800.956	-5.654	0.237	-5.654	8.79E-128	1.66E-126	CHST13
6H-0H	ENSG00000187037	372.343	4.239	0.162	4.239	7.52E-153	1.69E-151	GPR141
6H-0H	ENSG00000172548	698.034	4.837	0.235	4.837	1.22E-96	1.73E-95	NIPAL4
6H-0H	ENSG00000272666	168.659	4.150	0.208	4.150	1.32E-90	1.80E-89	U62317.1
6H-0H	ENSG00000112299	289.433	4.543	0.171	4.543	7.86E-158	1.84E-156	VNN1
6H-0H	ENSG00000163734	194750.698	4.166	0.140	4.166	6.15E-196	1.84E-194	CXCL3
6H-0H	ENSG00000050730	26471.684	6.467	0.288	6.467	1.11E-113	1.85E-112	TNIP3
6H-0H	ENSG00000136826	615.415	4.020	0.130	4.020	6.08E-211	2.02E-209	KLF4
6H-0H	ENSG00000142224	329.923	4.231	0.179	4.231	1.10E-125	2.05E-124	IL19
6H-0H	ENSG00000225889	213.753	5.102	0.220	5.102	1.16E-121	2.06E-120	AC012368.1
6H-0H	ENSG00000154451	14028.408	6.466	0.215	6.466	6.72E-201	2.06E-199	GBP5
6H-0H	ENSG00000164932	152.364	4.299	0.222	4.299	1.65E-85	2.18E-84	CTHRC1
6H-0H	ENSG00000111012	1722.424	5.572	0.228	5.572	1.12E-133	2.19E-132	CYP27B1
6H-0H	ENSG00000187116	2165.762	4.445	0.134	4.445	5.64E-242	2.20E-240	LILRA5
6H-0H	ENSG00000169429	957877.722	5.463	0.157	5.463	5.00E-266	2.21E-264	CXCL8
6H-0H	ENSG00000235385	52.324	4.630	0.407	4.630	3.92E-32	2.22E-31	LINC02154
6H-0H	ENSG00000104112	79.928	4.951	0.326	4.951	2.65E-54	2.32E-53	SCG3
6H-0H	ENSG00000118557	864.205	-4.390	0.124	-4.390	5.07E-276	2.35E-274	PMFBP1
6H-0H	ENSG00000105976	496.717	5.293	0.287	5.293	1.93E-78	2.37E-77	MET
6H-0H	ENSG00000041982	410.129	5.482	0.271	5.482	1.78E-93	2.47E-92	TNC
6H-0H	ENSG00000155011	211.328	4.025	0.252	4.025	2.70E-59	2.54E-58	DKK2
6H-0H	ENSG00000130589	3242.348	4.185	0.142	4.185	8.74E-193	2.58E-191	HELZ2
6H-0H	ENSG00000159166	207.451	5.034	0.208	5.034	1.40E-131	2.71E-130	LAD1
6H-0H	ENSG00000243509	199.403	4.308	0.197	4.308	1.71E-108	2.72E-107	TNFRSF6B
6H-0H	ENSG00000125726	622.218	4.085	0.112	4.085	5.41E-294	2.72E-292	CD70
6H-0H	ENSG00000224846	101.144	4.641	0.257	4.641	2.39E-75	2.81E-74	NQO2-AS1
6H-0H	ENSG00000160285	2409.016	4.279	0.155	4.279	1.12E-170	2.83E-169	LSS
6H-0H	ENSG00000171659	538.025	-6.211	0.416	-6.211	3.32E-52	2.84E-51	GPR34
6H-0H	ENSG00000160326	13139.099	4.463	0.156	4.463	1.04E-182	2.88E-181	SLC2A6
6H-0H	ENSG00000274736	1033.644	5.704	0.155	5.704	5.60E-300	2.93E-298	CCL23
6H-0H	ENSG00000134042	905.119	-4.271	0.170	-4.271	1.44E-141	2.98E-140	MRO
6H-0H	ENSG00000260997	173.849	4.916	0.211	4.916	1.73E-122	3.09E-121	AC004847.1
6H-0H	ENSG00000279400	154.359	4.244	0.251	4.244	3.02E-66	3.14E-65	AC008957.3
6H-0H	ENSG00000161640	189.949	-5.008	0.324	-5.008	3.58E-56	3.20E-55	SIGLEC11

6H-0H	ENSG00000188820	4323.395	4.250	0.145	4.250	1.15E-190	3.37E-189	CALHM6
6H-0H	ENSG00000111331	12473.593	4.312	0.218	4.312	2.49E-89	3.37E-88	OAS3
6H-0H	ENSG00000106341	101.289	4.670	0.245	4.670	2.72E-83	3.48E-82	PPP1R17
6H-0H	ENSG00000171227	118.075	-5.192	0.410	-5.192	5.43E-39	3.61E-38	TMEM37
6H-0H	ENSG00000110944	773.865	4.572	0.164	4.572	1.46E-172	3.72E-171	IL23A
6H-0H	ENSG00000253320	4155.495	4.438	0.145	4.438	1.16E-206	3.73E-205	AZIN1-AS1
6H-0H	ENSG00000172602	503.240	4.397	0.258	4.397	3.59E-67	3.78E-66	RND1
6H-0H	ENSG00000166016	569.228	5.590	0.176	5.590	1.09E-222	3.79E-221	ABTB2
6H-0H	ENSG00000110492	111.787	4.064	0.214	4.064	3.03E-82	3.85E-81	MDK
6H-0H	ENSG00000213949	235.927	4.619	0.173	4.619	1.67E-158	3.92E-157	ITGA1
6H-0H	ENSG00000229419	222.560	4.052	0.266	4.052	4.61E-54	4.02E-53	RALGAPA1P1
6H-0H	ENSG00000266709	143.733	5.210	0.376	5.210	5.58E-46	4.25E-45	AC005224.3
6H-0H	ENSG00000136048	11281.954	4.026	0.140	4.026	1.63E-182	4.53E-181	DRAM1
6H-0H	ENSG00000109756	9118.414	4.686	0.128	4.686	8.91E-298	4.57E-296	RAPGEF2
6H-0H	ENSG00000122641	12921.064	8.853	0.368	8.853	2.44E-129	4.66E-128	INHBA
6H-0H	ENSG00000162444	428.611	-4.549	0.265	-4.549	4.51E-68	4.81E-67	RBP7
6H-0H	ENSG00000078401	387.575	4.865	0.196	4.865	2.45E-138	4.94E-137	EDN1
6H-0H	ENSG00000163606	438.380	-4.532	0.194	-4.532	2.84E-122	5.06E-121	CD200R1
6H-0H	ENSG00000125148	24312.271	4.639	0.138	4.639	1.26E-250	5.11E-249	MT2A
6H-0H	ENSG00000124256	444.887	4.195	0.129	4.195	1.40E-235	5.27E-234	ZBP1
6H-0H	ENSG00000166527	341.592	4.470	0.164	4.470	2.19E-166	5.41E-165	CLEC4D
6H-0H	ENSG00000118503	47384.517	4.798	0.131	4.798	1.09E-297	5.59E-296	TNFAIP3
6H-0H	ENSG00000140511	816.261	6.817	0.191	6.817	1.19E-279	5.65E-278	HAPLN3
6H-0H	ENSG00000162433	4541.475	5.077	0.274	5.077	4.61E-79	5.69E-78	AK4
6H-0H	ENSG00000185245	426.432	4.205	0.162	4.205	2.61E-151	5.79E-150	GP1BA
6H-0H	ENSG00000131459	175.645	5.413	0.274	5.413	4.55E-89	6.14E-88	GFPT2
6H-0H	ENSG00000242048	1944.856	4.927	0.261	4.927	5.28E-82	6.69E-81	AC093583.1
6H-0H	ENSG00000105711	245.741	5.851	0.265	5.851	4.54E-110	7.35E-109	SCN1B
6H-0H	ENSG00000251136	1018.147	4.009	0.181	4.009	4.72E-110	7.64E-109	AF117829.1
6H-0H	ENSG00000075651	2117.322	4.043	0.119	4.043	2.02E-254	8.39E-253	PLD1
6H-0H	ENSG00000261040	473.287	4.495	0.276	4.495	8.82E-62	8.51E-61	WFDC21P
6H-0H	ENSG00000198133	250.507	4.187	0.180	4.187	4.82E-122	8.55E-121	TMEM229B
6H-0H	ENSG00000162614	704.251	4.020	0.113	4.020	1.88E-278	8.85E-277	NEXN
6H-0H	ENSG00000169220	495.764	-4.315	0.163	-4.315	4.08E-156	9.45E-155	RGS14
6H-0H	ENSG00000125430	9668.521	6.755	0.221	6.755	3.08E-206	9.85E-205	HS3ST3B1
6H-6H_control	ENSG00000143110	10408.043	-4.248	0.107	-4.248	0.000	0.000	C1orf162
6H-6H_control	ENSG00000168389	5675.495	4.035	0.096	4.035	0.000	0.000	MFSD2A
6H-6H_control	ENSG00000136514	1967.192	4.043	0.081	4.043	0.000	0.000	RTP4

6H-6H_control	ENSG00000117525	8301.128	4.056	0.094	4.056	0.000	0.000	F3
6H-6H_control	ENSG00000103569	24194.126	4.062	0.107	4.062	0.000	0.000	AQP9
6H-6H_control	ENSG00000177409	12986.735	4.073	0.087	4.073	0.000	0.000	SAMD9L
6H-6H_control	ENSG00000119917	130253.833	4.175	0.069	4.175	0.000	0.000	IFIT3
6H-6H_control	ENSG00000117226	2634.389	4.177	0.087	4.177	0.000	0.000	GBP3
6H-6H_control	ENSG00000136689	96893.244	4.186	0.061	4.186	0.000	0.000	IL1RN
6H-6H_control	ENSG00000135604	11094.992	4.191	0.098	4.191	0.000	0.000	STX11
6H-6H_control	ENSG00000171522	5563.745	4.196	0.106	4.196	0.000	0.000	PTGER4
6H-6H_control	ENSG00000138642	5878.802	4.198	0.064	4.198	0.000	0.000	HERC6
6H-6H_control	ENSG00000197536	1284.453	4.212	0.099	4.212	0.000	0.000	IRF1-AS1
6H-6H_control	ENSG00000139597	3079.234	4.245	0.082	4.245	0.000	0.000	N4BP2L1
6H-6H_control	ENSG00000165949	1577.870	4.339	0.099	4.339	0.000	0.000	IFI27
6H-6H_control	ENSG00000149798	2236.117	4.357	0.106	4.357	0.000	0.000	CDC42EP2
6H-6H_control	ENSG00000108771	4894.264	4.357	0.067	4.357	0.000	0.000	DHX58
6H-6H_control	ENSG00000197329	21369.894	4.379	0.065	4.379	0.000	0.000	PELI1
6H-6H_control	ENSG00000135842	8274.161	4.384	0.101	4.384	0.000	0.000	NIBAN1
6H-6H_control	ENSG00000140379	28405.208	4.435	0.082	4.435	0.000	0.000	BCL2A1
6H-6H_control	ENSG00000215861	9655.444	4.443	0.115	4.443	0.000	0.000	AC245297.1
6H-6H_control	ENSG00000056972	1183.501	4.507	0.100	4.507	0.000	0.000	TRAF3IP2
6H-6H_control	ENSG00000100342	2951.766	4.549	0.083	4.549	0.000	0.000	APOL1
6H-6H_control	ENSG00000134326	5082.369	4.584	0.117	4.584	0.000	0.000	CMPK2
6H-6H_control	ENSG00000099860	14927.309	4.598	0.063	4.598	0.000	0.000	GADD45B
6H-6H_control	ENSG00000135114	44581.118	4.625	0.054	4.625	0.000	0.000	OASL
6H-6H_control	ENSG00000184979	9987.071	4.744	0.088	4.744	0.000	0.000	USP18
6H-6H_control	ENSG00000086300	13446.577	4.830	0.097	4.830	0.000	0.000	SNX10
6H-6H_control	ENSG00000138646	18520.874	4.838	0.098	4.838	0.000	0.000	HERC5
6H-6H_control	ENSG00000123689	55459.349	4.864	0.100	4.864	0.000	0.000	G0S2

6H-6H_control	ENSG00000101187	1420.503	4.864	0.109	4.864	0.000	0.000	SLCO4A1
6H-6H_control	ENSG00000162645	17760.097	4.912	0.070	4.912	0.000	0.000	GBP2
6H-6H_control	ENSG00000109756	9118.414	4.922	0.127	4.922	0.000	0.000	RAPGEF2
6H-6H_control	ENSG00000151726	49643.687	4.937	0.098	4.937	0.000	0.000	ACSL1
6H-6H_control	ENSG00000185885	4903.850	4.946	0.126	4.946	0.000	0.000	IFITM1
6H-6H_control	ENSG00000125726	622.218	4.952	0.126	4.952	0.000	0.000	CD70
6H-6H_control	ENSG00000187608	187462.202	4.958	0.092	4.958	0.000	0.000	ISG15
6H-6H_control	ENSG00000285441	16399.973	5.038	0.107	5.038	0.000	0.000	SOD2
6H-6H_control	ENSG00000184588	6720.511	5.047	0.061	5.047	0.000	0.000	PDE4B
6H-6H_control	ENSG00000010818	2024.026	5.056	0.114	5.056	0.000	0.000	HIVEP2
6H-6H_control	ENSG00000104432	1139.914	5.104	0.109	5.104	0.000	0.000	IL7
6H-6H_control	ENSG00000121858	10453.628	5.154	0.089	5.154	0.000	0.000	TNFSF10
6H-6H_control	ENSG00000128284	3275.575	5.156	0.065	5.156	0.000	0.000	APOL3
6H-6H_control	ENSG00000169245	119889.974	5.210	0.117	5.210	0.000	0.000	CXCL10
6H-6H_control	ENSG00000136960	779.780	5.210	0.116	5.210	0.000	0.000	ENPP2
6H-6H_control	ENSG00000106701	4950.809	5.245	0.095	5.245	0.000	0.000	FSD1L
6H-6H_control	ENSG00000102524	29275.856	5.363	0.074	5.363	0.000	0.000	TNFSF13B
6H-6H_control	ENSG00000174837	789.252	5.435	0.118	5.435	0.000	0.000	ADGRE1
6H-6H_control	ENSG00000137393	10371.174	5.475	0.098	5.475	0.000	0.000	RNF144B
6H-6H_control	ENSG00000049249	4207.115	5.514	0.134	5.514	0.000	0.000	TNFRSF9
6H-6H_control	ENSG00000101460	2670.733	5.529	0.101	5.529	0.000	0.000	MAP1LC3A
6H-6H_control	ENSG00000170542	38582.781	5.535	0.101	5.535	0.000	0.000	SERPINB9
6H-6H_control	ENSG00000117228	47436.532	5.544	0.076	5.544	0.000	0.000	GBP1
6H-6H_control	ENSG00000145287	2413.342	5.546	0.101	5.546	0.000	0.000	PLAC8
6H-6H_control	ENSG00000168685	28002.502	5.567	0.106	5.567	0.000	0.000	IL7R
6H-6H_control	ENSG00000164136	2620.759	5.614	0.102	5.614	0.000	0.000	IL15
6H-6H_control	ENSG00000128271	6503.071	5.620	0.098	5.620	0.000	0.000	ADORA2A

6H-6H_control	ENSG00000105855	11817.836	5.635	0.130	5.635	0.000	0.000	ITGB8
6H-6H_control	ENSG00000182541	1590.193	5.725	0.094	5.725	0.000	0.000	LIMK2
6H-6H_control	ENSG00000134321	93857.564	5.993	0.120	5.993	0.000	0.000	RSAD2
6H-6H_control	ENSG00000274736	1033.644	6.007	0.155	6.007	0.000	0.000	CCL23
6H-6H_control	ENSG00000169429	957877.722	6.009	0.157	6.009	0.000	0.000	CXCL8
6H-6H_control	ENSG00000171049	1916.493	6.054	0.131	6.054	0.000	0.000	FPR2
6H-6H_control	ENSG00000117090	1752.672	6.120	0.118	6.120	0.000	0.000	SLAMF1
6H-6H_control	ENSG00000128383	22244.347	6.126	0.120	6.126	0.000	0.000	APOBEC3A
6H-6H_control	ENSG00000123700	8458.045	6.203	0.074	6.203	0.000	0.000	KCNJ2
6H-6H_control	ENSG00000234883	24285.234	6.218	0.098	6.218	0.000	0.000	MIR155HG
6H-6H_control	ENSG00000138378	5160.835	6.234	0.141	6.234	0.000	0.000	STAT4
6H-6H_control	ENSG00000131979	1421.850	6.342	0.145	6.342	0.000	0.000	GCH1
6H-6H_control	ENSG00000186827	5014.738	6.431	0.161	6.431	0.000	0.000	TNFRSF4
6H-6H_control	ENSG00000089351	5267.991	6.530	0.122	6.530	0.000	0.000	GRAMD1A
6H-6H_control	ENSG00000010030	1392.680	6.551	0.131	6.551	0.000	0.000	ETV7
6H-6H_control	ENSG00000073756	21640.883	6.573	0.151	6.573	0.000	0.000	PTGS2
6H-6H_control	ENSG00000115008	4893.793	6.631	0.143	6.631	0.000	0.000	IL1A
6H-6H_control	ENSG00000140511	816.261	6.669	0.172	6.669	0.000	0.000	HAPLN3
6H-6H_control	ENSG00000004468	1997.236	6.716	0.113	6.716	0.000	0.000	CD38
6H-6H_control	ENSG00000134470	7645.099	6.772	0.093	6.772	0.000	0.000	IL15RA
6H-6H_control	ENSG00000137757	982.758	6.853	0.163	6.853	0.000	0.000	CASP5
6H-6H_control	ENSG00000253831	889.036	6.885	0.172	6.885	0.000	0.000	ETV3L
6H-6H_control	ENSG00000125538	406650.854	6.894	0.100	6.894	0.000	0.000	IL1B
6H-6H_control	ENSG00000172183	34650.706	6.959	0.078	6.959	0.000	0.000	ISG20
6H-6H_control	ENSG00000026751	95133.604	7.018	0.102	7.018	0.000	0.000	SLAMF7
6H-6H_control	ENSG00000255398	4595.526	7.086	0.161	7.086	0.000	0.000	HCAR3
6H-6H_control	ENSG00000205755	1248.343	7.173	0.172	7.173	0.000	0.000	CRLF2

6H-6H_control	ENSG00000277632	681845.152	7.197	0.100	7.197	0.000	0.000	CCL3
6H-6H_control	ENSG00000120217	11894.744	7.215	0.117	7.215	0.000	0.000	CD274
6H-6H_control	ENSG00000078081	7768.475	7.278	0.107	7.278	0.000	0.000	LAMP3
6H-6H_control	ENSG00000162654	16541.234	7.303	0.088	7.303	0.000	0.000	GBP4
6H-6H_control	ENSG00000121594	3131.334	7.380	0.146	7.380	0.000	0.000	CD80
6H-6H_control	ENSG00000105246	4390.908	7.402	0.137	7.402	0.000	0.000	EBI3
6H-6H_control	ENSG00000134460	2344.532	7.491	0.126	7.491	0.000	0.000	IL2RA
6H-6H_control	ENSG00000153898	3133.356	7.545	0.115	7.545	0.000	0.000	MCOLN2
6H-6H_control	ENSG00000166165	3861.304	7.618	0.133	7.618	0.000	0.000	CKB
6H-6H_control	ENSG00000276085	243846.112	7.622	0.090	7.622	0.000	0.000	CCL3L1
6H-6H_control	ENSG00000120436	1755.796	7.786	0.156	7.786	0.000	0.000	GPR31
6H-6H_control	ENSG00000108700	99365.331	7.818	0.143	7.818	0.000	0.000	CCL8
6H-6H_control	ENSG00000169248	7858.924	7.851	0.151	7.851	0.000	0.000	CXCL11
6H-6H_control	ENSG00000115009	61295.772	7.854	0.148	7.854	0.000	0.000	CCL20
6H-6H_control	ENSG00000197272	6763.347	7.879	0.134	7.879	0.000	0.000	IL27
6H-6H_control	ENSG00000271503	137138.905	7.937	0.063	7.937	0.000	0.000	CCL5
6H-6H_control	ENSG00000276070	353729.863	8.132	0.105	8.132	0.000	0.000	CCL4L2
6H-6H_control	ENSG00000275302	461015.420	8.133	0.091	8.133	0.000	0.000	CCL4
6H-6H_control	ENSG00000104213	1888.722	8.480	0.202	8.480	0.000	0.000	PDGFRL
6H-6H_control	ENSG00000243649	4967.446	8.947	0.162	8.947	0.000	0.000	CFB
6H-6H_control	ENSG00000251230	15100.069	9.144	0.228	9.144	0.000	0.000	MIR3945HG
6H-6H_control	ENSG00000123610	64667.055	9.276	0.105	9.276	0.000	0.000	TNFAIP6
6H-6H_control	ENSG00000136244	17748.154	9.340	0.148	9.340	0.000	0.000	IL6
6H-6H_control	ENSG00000131203	64101.670	9.738	0.069	9.738	0.000	0.000	IDO1
6H-6H_control	ENSG00000102794	23702.223	9.776	0.107	9.776	0.000	0.000	ACOD1
6H-6H_control	ENSG00000148677	644.817	4.097	0.258	4.097	1.23E-58	1.01E-57	ANKRD1
6H-6H_control	ENSG00000134042	905.119	-4.742	0.169	-4.742	4.30E-175	1.04E-173	MRO

6H-6H_control	ENSG00000111012	1722.424	6.467	0.233	6.467	4.37E-172	1.04E-170	CYP27B1
6H-6H_control	ENSG00000124875	3228.416	5.036	0.147	5.036	2.79E-260	1.06E-258	CXCL6
6H-6H_control	ENSG00000159166	207.451	5.809	0.230	5.809	5.27E-143	1.06E-141	LAD1
6H-6H_control	ENSG00000251136	1018.147	4.432	0.183	4.432	5.88E-131	1.09E-129	AF117829.1
6H-6H_control	ENSG00000163661	2000.863	5.237	0.144	5.237	2.41E-293	1.09E-291	PTX3
6H-6H_control	ENSG00000166527	341.592	5.585	0.191	5.585	4.14E-191	1.10E-189	CLEC4D
6H-6H_control	ENSG00000169220	495.764	-4.632	0.162	-4.632	4.48E-181	1.12E-179	RGS14
6H-6H_control	ENSG00000172548	698.034	4.957	0.224	4.957	7.61E-111	1.15E-109	NIPAL4
6H-6H_control	ENSG00000163734	194750.698	4.612	0.140	4.612	3.56E-239	1.25E-237	CXCL3
6H-6H_control	ENSG00000197646	416.516	4.585	0.160	4.585	5.25E-183	1.32E-181	PDCD1LG2
6H-6H_control	ENSG00000078401	387.575	4.283	0.160	4.283	5.91E-160	1.33E-158	EDN1
6H-6H_control	ENSG00000282608	1265.186	-7.259	0.250	-7.259	5.27E-187	1.36E-185	ADORA3
6H-6H_control	ENSG00000272666	168.659	4.122	0.197	4.122	1.02E-98	1.37E-97	U62317.1
6H-6H_control	ENSG00000224846	101.144	4.600	0.239	4.600	1.27E-84	1.45E-83	NQO2-AS1
6H-6H_control	ENSG00000205856	266.648	4.096	0.179	4.096	9.00E-118	1.46E-116	C22orf42
6H-6H_control	ENSG00000115267	12591.394	4.131	0.234	4.131	1.50E-71	1.47E-70	IFIH1
6H-6H_control	ENSG00000162444	428.611	-4.453	0.265	-4.453	1.64E-65	1.48E-64	RBP7
6H-6H_control	ENSG00000163739	247998.451	5.936	0.162	5.936	3.26E-297	1.51E-295	CXCL1
6H-6H_control	ENSG00000198133	250.507	4.214	0.169	4.214	7.74E-139	1.53E-137	TMEM229B
6H-6H_control	ENSG00000266709	143.733	5.932	0.421	5.932	2.26E-47	1.56E-46	AC005224.3
6H-6H_control	ENSG00000182782	3118.059	7.695	0.212	7.695	3.62E-289	1.62E-287	HCAR2
6H-6H_control	ENSG00000092068	8254.861	-6.720	0.230	-6.720	6.56E-189	1.72E-187	SLC7A8
6H-6H_control	ENSG00000172602	503.240	5.173	0.299	5.173	1.80E-69	1.73E-68	RND1
6H-6H_control	ENSG00000197093	942.311	-5.592	0.179	-5.592	5.90E-216	1.81E-214	GAL3ST4
6H-6H_control	ENSG00000110944	773.865	5.353	0.181	5.353	6.59E-195	1.82E-193	IL23A
6H-6H_control	ENSG00000082781	435.134	-4.005	0.180	-4.005	1.20E-111	1.83E-110	ITGB5
6H-6H_control	ENSG00000243509	199.403	4.847	0.203	4.847	1.02E-127	1.84E-126	TNFRSF6B

6H-6H_control	ENSG00000119865	318.495	-4.093	0.204	-4.093	1.49E-91	1.87E-90	CNRIP1
6H-6H_control	ENSG00000184489	112.024	4.567	0.244	4.567	1.87E-80	2.03E-79	PTP4A3
6H-6H_control	ENSG00000154451	14028.408	6.647	0.214	6.647	6.77E-213	2.04E-211	GBP5
6H-6H_control	ENSG00000235385	52.324	4.636	0.381	4.636	3.76E-36	2.07E-35	LINC02154
6H-6H_control	ENSG00000188820	4323.395	4.363	0.145	4.363	8.05E-202	2.28E-200	CALHM6
6H-6H_control	ENSG00000164932	152.364	4.061	0.197	4.061	1.74E-96	2.32E-95	CTHRC1
6H-6H_control	ENSG00000122641	12921.064	10.042	0.377	10.042	1.05E-157	2.33E-156	INHBA
6H-6H_control	ENSG00000260997	173.849	4.608	0.182	4.608	1.19E-143	2.45E-142	AC004847.1
6H-6H_control	ENSG00000105711	245.741	5.927	0.252	5.927	1.42E-124	2.51E-123	SCN1B
6H-6H_control	ENSG00000171659	538.025	-7.089	0.414	-7.089	2.73E-67	2.53E-66	GPR34
6H-6H_control	ENSG00000167772	61.661	4.150	0.364	4.150	5.12E-32	2.56E-31	ANGPTL4
6H-6H_control	ENSG00000165474	3489.483	8.718	0.251	8.718	6.62E-266	2.62E-264	GJB2
6H-6H_control	ENSG00000105825	686.225	4.612	0.185	4.612	1.33E-138	2.63E-137	TFPI2
6H-6H_control	ENSG00000160326	13139.099	4.704	0.156	4.704	9.68E-203	2.77E-201	SLC2A6
6H-6H_control	ENSG00000146278	14296.318	4.078	0.239	4.078	3.04E-67	2.81E-66	PNRC1
6H-6H_control	ENSG00000185245	426.432	5.174	0.184	5.174	1.18E-175	2.87E-174	GP1BA
6H-6H_control	ENSG00000124256	444.887	4.470	0.129	4.470	7.27E-267	2.91E-265	ZBP1
6H-6H_control	ENSG00000123240	3893.161	4.103	0.138	4.103	1.17E-196	3.24E-195	OPTN
6H-6H_control	ENSG00000156234	86.881	6.943	0.409	6.943	3.56E-66	3.25E-65	CXCL13
6H-6H_control	ENSG00000136826	615.415	4.217	0.128	4.217	9.33E-240	3.30E-238	KLF4
6H-6H_control	ENSG00000118557	864.205	-4.020	0.124	-4.020	9.86E-232	3.35E-230	PMFBP1
6H-6H_control	ENSG00000213949	235.927	4.793	0.169	4.793	1.40E-178	3.44E-177	ITGA1
6H-6H_control	ENSG00000105976	496.717	5.090	0.278	5.090	3.29E-77	3.45E-76	MET
6H-6H_control	ENSG00000183160	352.198	-4.346	0.170	-4.346	1.68E-145	3.50E-144	TMEM119
6H-6H_control	ENSG00000124216	219.320	5.647	0.255	5.647	2.32E-111	3.52E-110	SNAI1
6H-6H_control	ENSG00000280228	104.630	4.145	0.378	4.145	7.54E-30	3.57E-29	AC079753.1
6H-6H_control	ENSG00000173638	304.127	-4.136	0.187	-4.136	2.65E-110	3.96E-109	SLC19A1

6H-6H_control	ENSG00000118503	47384.517	4.805	0.130	4.805	8.61E-300	4.12E-298	TNFAIP3
6H-6H_control	ENSG00000184557	10409.754	6.755	0.276	6.755	2.26E-134	4.29E-133	SOCS3
6H-6H_control	ENSG00000115919	69786.626	4.350	0.137	4.350	1.41E-223	4.50E-222	KYNU
6H-6H_control	ENSG00000198719	220.149	4.454	0.211	4.454	3.38E-101	4.66E-100	DLL1
6H-6H_control	ENSG00000131459	175.645	5.968	0.299	5.968	3.82E-91	4.75E-90	GFPT2
6H-6H_control	ENSG00000136048	11281.954	4.542	0.141	4.542	1.43E-230	4.77E-229	DRAM1
6H-6H_control	ENSG00000229419	222.560	4.238	0.264	4.238	5.71E-60	4.79E-59	RALGAPA1P1
6H-6H_control	ENSG00000163606	438.380	-4.274	0.194	-4.274	3.32E-109	4.91E-108	CD200R1
6H-6H_control	ENSG00000050730	26471.684	6.857	0.288	6.857	2.85E-127	5.10E-126	TNIP3
6H-6H_control	ENSG00000253593	91.357	4.088	0.261	4.088	6.63E-57	5.26E-56	AC110741.1
6H-6H_control	ENSG00000112303	600.318	4.712	0.126	4.712	1.08E-307	5.33E-306	VNN2
6H-6H_control	ENSG00000104112	79.928	4.052	0.254	4.052	6.54E-59	5.38E-58	SCG3
6H-6H_control	ENSG00000074660	441.135	4.012	0.320	4.012	1.00E-37	5.70E-37	SCARF1
6H-6H_control	ENSG00000166670	1616.387	7.972	0.245	7.972	1.74E-233	5.92E-232	MMP10
6H-6H_control	ENSG00000225889	213.753	5.498	0.227	5.498	3.20E-132	5.98E-131	AC012368.1
6H-6H_control	ENSG00000166920	105742.630	4.465	0.197	4.465	3.88E-116	6.18E-115	C15orf48
6H-6H_control	ENSG00000180316	495.415	4.454	0.181	4.454	3.23E-136	6.27E-135	PNPLA1
6H-6H_control	ENSG00000099282	41.324	-4.205	0.571	-4.205	2.11E-15	6.35E-15	TSPAN15
6H-6H_control	ENSG00000166523	10612.295	4.976	0.154	4.976	1.91E-231	6.47E-230	CLEC4E
6H-6H_control	ENSG00000150938	1524.969	4.596	0.156	4.596	2.43E-193	6.61E-192	CRIM1
6H-6H_control	ENSG00000180767	800.956	-6.214	0.237	-6.214	3.06E-154	6.69E-153	CHST13
6H-6H_control	ENSG00000162433	4541.475	5.434	0.274	5.434	5.64E-90	6.91E-89	AK4
6H-6H_control	ENSG00000106341	101.289	5.335	0.269	5.335	5.65E-90	6.91E-89	PPP1R17
6H-6H_control	ENSG00000120129	12153.273	4.407	0.557	4.407	2.28E-17	7.39E-17	DUSP1
6H-6H_control	ENSG00000224083	69.989	4.032	0.402	4.032	1.76E-25	7.48E-25	MTCO1P11
6H-6H_control	ENSG00000171227	118.075	-5.503	0.409	-5.503	1.17E-43	7.52E-43	TMEM37
6H-6H_control	ENSG00000152760	369.471	-4.739	0.185	-4.739	3.65E-147	7.68E-146	TCTEX1D1

6H-6H_control	ENSG00000187116	2165.762	4.373	0.133	4.373	2.21E-239	7.80E-238	LILRA5
6H-6H_control	ENSG00000253320	4155.495	4.758	0.146	4.758	2.31E-235	7.95E-234	AZIN1-AS1
6H-6H_control	ENSG00000242048	1944.856	5.566	0.263	5.566	5.72E-102	7.95E-101	AC093583.1
6H-6H_control	ENSG00000166016	569.228	5.214	0.149	5.214	1.94E-272	8.00E-271	ABTB2
6H-6H_control	ENSG00000140092	224.030	-4.205	0.221	-4.205	7.15E-83	8.01E-82	FBLN5
6H-6H_control	ENSG00000162599	77.649	-4.645	0.406	-4.645	1.66E-32	8.41E-32	NFIA
6H-6H_control	ENSG00000256262	335.685	4.290	0.137	4.290	2.78E-216	8.58E-215	USP30-AS1
6H-6H_control	ENSG00000269713	1482.714	4.065	0.115	4.065	2.07E-274	8.60E-273	NBPF9
6H-6H_control	ENSG00000110492	111.787	4.494	0.221	4.494	6.87E-94	8.84E-93	MDK
6H-6H_control	ENSG00000112299	289.433	4.058	0.153	4.058	4.03E-156	8.90E-155	VNN1
6H-6H_control	ENSG00000142224	329.923	4.933	0.189	4.933	4.25E-153	9.23E-152	IL19
6H-6H_control	ENSG00000125430	9668.521	6.624	0.220	6.624	3.28E-201	9.26E-200	HS3ST3B1
6H-6H_control	ENSG00000161640	189.949	-4.635	0.324	-4.635	1.37E-48	9.63E-48	SIGLEC11
6H-6H_control	ENSG00000250274	274.190	4.033	0.220	4.033	9.34E-77	9.74E-76	AC034199.1

Table 6 G:Profiler analysis of differentially aged gene clusters

source	term_name	term_id	term_size	effective_domain_size	adjusted_p_value__cluster1	query_size__cluster1	intersection_size_cluster1	adjusted_p_value__cluster2	query_size__cluster2	intersection_size_cluster2	adjusted_p_value__cluster3	query_size__cluster3	intersection_size_cluster3	adjusted_p_value__cluster4	query_size__cluster4	intersection_size_cluster4
TF	CDX-1	TF:M02086_1	5116	12028	1.06E-02	75	54	1.00E+00	88	49	1.00E+00	53	25	1.00E+00	44	18
TF	SOX-4	TF:M03849	7734	12028	1.32E-03	75	69	1.00E+00	88	68	1.00E+00	53	34	1.00E+00	44	33
TF	ZF5	TF:M00716	9887	12028	1.00E+00	75	70	1.50E-03	88	88	1.00E+00	53	49	1.00E+00	44	42
MIRNA	hsa-miR-106a-5p	MIRNA:hsa-miR-106a-5p	564	12028	2.37E-03	75	17	1.00E+00	88	11	1.00E+00	53	8	1.00E+00	44	1
MIRNA	hsa-miR-1234-3p	MIRNA:hsa-miR-1234-3p	118	12028	3.30E-02	75	8	1.00E+00	88	4	1.00E+00	53	2	1.00E+00	44	2
MIRNA	hsa-miR-1247-3p	MIRNA:hsa-miR-1247-3p	249	12028	1.90E-04	75	13	1.00E+00	88	8	1.00E+00	53	5	1.00E+00	44	6
MIRNA	hsa-miR-1273e	MIRNA:hsa-miR-1273e	332	12028	1.00E+00	75	7	5.76E-03	88	14	1.00E+00	53	5	1.00E+00	44	8
MIRNA	hsa-miR-1273h-5p	MIRNA:hsa-miR-1273h-5p	472	12028	5.27E-02	75	14	1.14E-09	88	24	1.00E+00	53	10	1.00E+00	44	9
MIRNA	hsa-miR-1304-3p	MIRNA:hsa-miR-1304-3p	366	12028	3.37E-06	75	17	5.71E-06	88	18	1.00E+00	53	9	1.00E+00	44	4
MIRNA	hsa-miR-1307-3p	MIRNA:hsa-miR-1307-3p	177	12028	6.84E-01	75	8	2.76E-04	88	12	1.00E+00	53	2	1.00E+00	-1	-1
MIRNA	hsa-miR-143-5p	MIRNA:hsa-miR-143-5p	301	12028	8.16E-01	75	10	1.30E-02	88	13	1.00E+00	53	1	1.00E+00	44	2
MIRNA	hsa-miR-149-3p	MIRNA:hsa-miR-149-3p	582	12028	6.01E-01	75	14	5.90E-06	88	22	1.00E+00	53	9	1.00E+00	44	7
MIRNA	hsa-miR-150-5p	MIRNA:hsa-miR-150-5p	381	12028	2.81E-02	75	13	3.08E-02	88	14	1.00E+00	53	9	1.00E+00	44	6
MIRNA	hsa-miR-1827	MIRNA:hsa-miR-1827	602	12028	1.00E+00	75	13	1.13E-05	88	22	1.00E+00	53	10	1.00E+00	44	7

MIRNA	hsa-miR-186-3p	MIRNA:hsa-miR-186-3p	317	12028	2.69E-02	75	12	2.35E-02	88	13	5.01E-02	53	10	1.00E+00	44	5
MIRNA	hsa-miR-24-3p	MIRNA:hsa-miR-24-3p	653	12028	1.00E+00	75	14	1.75E-03	88	20	1.00E+00	53	11	1.26E-01	44	12
MIRNA	hsa-miR-302a-3p	MIRNA:hsa-miR-302a-3p	352	12028	5.28E-01	75	11	1.00E+00	88	10	1.46E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-302b-3p	MIRNA:hsa-miR-302b-3p	349	12028	4.87E-01	75	11	1.00E+00	88	10	1.34E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-302c-3p	MIRNA:hsa-miR-302c-3p	385	12028	1.00E+00	75	11	2.06E-01	88	13	3.56E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-302d-3p	MIRNA:hsa-miR-302d-3p	350	12028	5.00E-01	75	11	1.00E+00	88	10	1.38E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-302e	MIRNA:hsa-miR-302e	340	12028	3.81E-01	75	11	1.00E+00	88	10	1.03E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-30b-3p	MIRNA:hsa-miR-30b-3p	477	12028	5.98E-02	75	14	1.44E-09	88	24	1.00E+00	53	10	1.00E+00	44	9
MIRNA	hsa-miR-30c-1-3p	MIRNA:hsa-miR-30c-1-3p	392	12028	2.48E-01	75	12	2.27E-07	88	20	1.00E+00	53	9	1.00E+00	44	7
MIRNA	hsa-miR-30c-2-3p	MIRNA:hsa-miR-30c-2-3p	401	12028	5.01E-02	75	13	3.44E-07	88	20	1.00E+00	53	8	1.00E+00	44	7
MIRNA	hsa-miR-3122	MIRNA:hsa-miR-3122	269	12028	4.57E-03	75	12	3.91E-07	88	17	1.00E+00	53	7	1.00E+00	44	7
MIRNA	hsa-miR-3135b	MIRNA:hsa-miR-3135b	294	12028	1.00E+00	75	9	1.27E-03	88	14	1.00E+00	53	3	1.00E+00	44	6
MIRNA	hsa-miR-3178	MIRNA:hsa-miR-3178	49	12028	1.00E+00	75	3	1.00E+00	-1	-1	1.00E+00	53	1	1.00E+00	44	1
MIRNA	hsa-miR-3194-3p	MIRNA:hsa-miR-3194-3p	104	12028	1.24E-02	75	8	1.00E+00	88	1	1.00E+00	53	4	1.00E+00	44	3
MIRNA	hsa-miR-340-3p	MIRNA:hsa-miR-340-3p	97	12028	1.00E+00	75	4	2.53E-02	88	8	1.00E+00	53	2	1.00E+00	44	1

MIRNA	hsa-miR-3652	MIRNA:hsa-miR-3652	298	12028	1.00E+00	75	9	1.50E-03	88	14	1.00E+00	53	3	1.00E+00	44	7
MIRNA	hsa-miR-3653-5p	MIRNA:hsa-miR-3653-5p	439	12028	1.00E+00	75	11	2.95E-02	88	15	1.00E+00	53	9	1.74E-02	44	11
MIRNA	hsa-miR-3663-5p	MIRNA:hsa-miR-3663-5p	155	12028	2.36E-02	75	9	1.00E+00	88	5	1.00E+00	53	1	1.00E+00	44	5
MIRNA	hsa-miR-3689a-3p	MIRNA:hsa-miR-3689a-3p	471	12028	5.14E-02	75	14	1.09E-09	88	24	1.00E+00	53	10	1.00E+00	44	9
MIRNA	hsa-miR-3689b-3p	MIRNA:hsa-miR-3689b-3p	471	12028	5.14E-02	75	14	1.09E-09	88	24	1.00E+00	53	10	1.00E+00	44	9
MIRNA	hsa-miR-3689c	MIRNA:hsa-miR-3689c	471	12028	5.14E-02	75	14	1.09E-09	88	24	1.00E+00	53	10	1.00E+00	44	9
MIRNA	hsa-miR-371a-5p	MIRNA:hsa-miR-371a-5p	240	12028	8.61E-01	75	9	1.00E+00	88	7	1.00E+00	53	4	1.00E+00	44	5
MIRNA	hsa-miR-372-3p	MIRNA:hsa-miR-372-3p	351	12028	5.14E-01	75	11	1.00E+00	88	10	1.42E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-377-5p	MIRNA:hsa-miR-377-5p	414	12028	1.00E+00	75	9	2.17E-03	88	16	1.00E+00	53	5	7.55E-01	44	9
MIRNA	hsa-miR-383-3p	MIRNA:hsa-miR-383-3p	308	12028	1.00E+00	75	9	1.69E-02	88	13	3.85E-02	53	10	1.00E+00	44	7
MIRNA	hsa-miR-3913-5p	MIRNA:hsa-miR-3913-5p	268	12028	4.39E-03	75	12	3.69E-07	88	17	1.00E+00	53	7	1.00E+00	44	7
MIRNA	hsa-miR-3934-5p	MIRNA:hsa-miR-3934-5p	194	12028	1.58E-02	75	10	1.00E+00	88	7	1.00E+00	53	5	1.00E+00	44	3
MIRNA	hsa-miR-4279	MIRNA:hsa-miR-4279	329	12028	1.00E+00	75	6	1.00E+00	88	10	1.00E+00	53	6	1.11E-02	44	10
MIRNA	hsa-miR-4284	MIRNA:hsa-miR-4284	465	12028	2.08E-07	75	20	6.44E-13	88	27	2.24E-01	53	11	3.16E-04	44	13
MIRNA	hsa-miR-4324	MIRNA:hsa-miR-4324	129	12028	1.00E+00	75	1	1.00E+00	88	4	1.00E+00	53	3	1.00E+00	44	1

MIRNA	hsa-miR-4430	MIRNA:hsa-miR-4430	298	12028	1.00E+00	75	9	1.50E-03	88	14	1.00E+00	53	3	1.00E+00	44	7
MIRNA	hsa-miR-4433a-3p	MIRNA:hsa-miR-4433a-3p	279	12028	1.00E+00	75	5	4.09E-02	88	12	1.00E+00	53	5	1.00E+00	44	4
MIRNA	hsa-miR-4438	MIRNA:hsa-miR-4438	239	12028	1.00E+00	75	6	1.00E+00	88	9	4.30E-02	53	9	1.00E+00	44	2
MIRNA	hsa-miR-4454	MIRNA:hsa-miR-4454	41	12028	1.00E+00	75	3	2.36E-02	88	6	1.00E+00	-1	-1	1.00E+00	-1	-1
MIRNA	hsa-miR-4459	MIRNA:hsa-miR-4459	507	12028	1.00E+00	75	11	3.14E-06	88	21	1.00E+00	53	9	5.51E-01	44	10
MIRNA	hsa-miR-4485-5p	MIRNA:hsa-miR-4485-5p	135	12028	1.00E+00	75	7	2.92E-02	88	9	1.00E+00	-1	-1	1.00E+00	44	1
MIRNA	hsa-miR-450a-1-3p	MIRNA:hsa-miR-450a-1-3p	282	12028	8.47E-04	75	13	8.30E-07	88	17	1.00E+00	53	7	1.00E+00	44	7
MIRNA	hsa-miR-4537	MIRNA:hsa-miR-4537	230	12028	6.13E-01	75	9	4.32E-02	88	11	1.00E+00	53	7	1.00E+00	44	3
MIRNA	hsa-miR-455-3p	MIRNA:hsa-miR-455-3p	590	12028	1.00E+00	75	11	1.09E-06	88	23	1.00E+00	53	9	1.00E+00	44	9
MIRNA	hsa-miR-4638-5p	MIRNA:hsa-miR-4638-5p	179	12028	1.00E+00	75	7	3.13E-04	88	12	1.00E+00	53	2	1.00E+00	44	1
MIRNA	hsa-miR-4684-5p	MIRNA:hsa-miR-4684-5p	216	12028	1.00E+00	75	8	2.58E-04	88	13	1.00E+00	53	5	1.00E+00	44	2
MIRNA	hsa-miR-4695-5p	MIRNA:hsa-miR-4695-5p	272	12028	1.00E+00	75	6	3.13E-02	88	12	1.00E+00	53	4	1.00E+00	44	4
MIRNA	hsa-miR-4716-5p	MIRNA:hsa-miR-4716-5p	127	12028	1.00E+00	75	5	1.36E-03	88	10	1.00E+00	53	4	1.00E+00	44	1
MIRNA	hsa-miR-4728-5p	MIRNA:hsa-miR-4728-5p	587	12028	6.62E-01	75	14	6.96E-06	88	22	1.00E+00	53	9	1.00E+00	44	7
MIRNA	hsa-miR-4731-5p	MIRNA:hsa-miR-4731-5p	281	12028	8.12E-04	75	13	3.02E-01	88	11	1.00E+00	53	7	1.00E+00	44	3

MIRNA	hsa-miR-4768-3p	MIRNA:hsa-miR-4768-3p	471	12028	1.00E+00	75	9	1.27E-02	88	16	2.54E-01	53	11	1.00E+00	44	6
MIRNA	hsa-miR-4772-3p	MIRNA:hsa-miR-4772-3p	231	12028	1.00E+00	75	8	9.89E-12	88	20	3.45E-01	53	8	1.00E+00	44	4
MIRNA	hsa-miR-4775	MIRNA:hsa-miR-4775	202	12028	2.29E-02	75	10	1.00E+00	88	8	1.00E+00	53	1	1.00E+00	44	2
MIRNA	hsa-miR-498	MIRNA:hsa-miR-498	222	12028	1.00E+00	75	7	3.59E-04	88	13	1.00E+00	53	3	7.12E-01	44	7
MIRNA	hsa-miR-504-3p	MIRNA:hsa-miR-504-3p	294	12028	1.00E+00	75	7	1.48E-04	88	15	1.00E+00	53	3	1.00E+00	44	6
MIRNA	hsa-miR-508-5p	MIRNA:hsa-miR-508-5p	332	12028	3.03E-01	75	11	5.76E-03	88	14	6.34E-01	53	9	1.00E+00	44	2
MIRNA	hsa-miR-5089-5p	MIRNA:hsa-miR-5089-5p	272	12028	3.37E-01	75	10	4.72E-04	88	14	1.00E+00	53	7	1.00E+00	44	3
MIRNA	hsa-miR-5095	MIRNA:hsa-miR-5095	307	12028	9.67E-01	75	10	1.63E-02	88	13	2.55E-05	53	13	1.00E+00	44	3
MIRNA	hsa-miR-515-5p	MIRNA:hsa-miR-515-5p	132	12028	1.00E+00	75	4	2.41E-02	88	9	1.00E+00	-1	-1	1.00E+00	44	4
MIRNA	hsa-miR-519e-5p	MIRNA:hsa-miR-519e-5p	132	12028	1.00E+00	75	4	2.41E-02	88	9	1.00E+00	-1	-1	1.00E+00	44	4
MIRNA	hsa-miR-520a-3p	MIRNA:hsa-miR-520a-3p	346	12028	4.49E-01	75	11	1.00E+00	88	10	1.23E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-520b	MIRNA:hsa-miR-520b	345	12028	4.37E-01	75	11	1.00E+00	88	10	1.19E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-520c-3p	MIRNA:hsa-miR-520c-3p	345	12028	4.37E-01	75	11	1.00E+00	88	10	1.19E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-520d-3p	MIRNA:hsa-miR-520d-3p	340	12028	3.81E-01	75	11	1.00E+00	88	10	1.03E-02	53	11	1.00E+00	44	2
MIRNA	hsa-miR-520e	MIRNA:hsa-miR-520e	342	12028	4.02E-01	75	11	1.00E+00	88	10	1.09E-02	53	11	1.00E+00	44	2

MIRNA	hsa-miR-520g-3p	MIRNA:hsa-miR-520g-3p	291	12028	8.51E-02	75	11	1.00E+00	88	10	2.29E-02	53	10	1.00E+00	44	3
MIRNA	hsa-miR-520h	MIRNA:hsa-miR-520h	293	12028	9.10E-02	75	11	1.00E+00	88	10	2.44E-02	53	10	1.00E+00	44	3
MIRNA	hsa-miR-526b-3p	MIRNA:hsa-miR-526b-3p	502	12028	1.86E-02	75	15	1.00E+00	88	11	1.00E+00	53	8	1.00E+00	44	1
MIRNA	hsa-miR-544b	MIRNA:hsa-miR-544b	155	12028	1.00E+00	-1	-1	1.00E+00	88	4	1.06E-03	53	9	1.00E+00	44	1
MIRNA	hsa-miR-548as-3p	MIRNA:hsa-miR-548as-3p	121	12028	1.00E+00	75	2	1.15E-02	88	9	1.00E+00	-1	-1	1.00E+00	44	1
MIRNA	hsa-miR-548s	MIRNA:hsa-miR-548s	313	12028	2.35E-02	75	12	2.77E-03	88	14	1.00E+00	53	5	1.00E+00	44	6
MIRNA	hsa-miR-5589-5p	MIRNA:hsa-miR-5589-5p	252	12028	2.20E-04	75	13	1.00E+00	88	9	1.00E+00	53	7	1.00E+00	44	2
MIRNA	hsa-miR-5693	MIRNA:hsa-miR-5693	261	12028	1.00E+00	75	9	2.87E-05	88	15	1.00E+00	53	2	1.00E+00	44	7
MIRNA	hsa-miR-5697	MIRNA:hsa-miR-5697	130	12028	1.00E+00	75	1	1.00E+00	88	7	1.00E+00	53	1	1.00E+00	44	1
MIRNA	hsa-miR-6086	MIRNA:hsa-miR-6086	396	12028	1.00E+00	75	9	1.17E-03	88	16	1.00E+00	53	5	5.29E-01	44	9
MIRNA	hsa-miR-6087	MIRNA:hsa-miR-6087	183	12028	1.00E+00	75	6	1.00E+00	88	6	1.00E+00	53	2	1.37E-02	44	8
MIRNA	hsa-miR-6089	MIRNA:hsa-miR-6089	120	12028	1.00E+00	75	1	1.07E-02	88	9	1.00E+00	53	3	1.00E+00	44	2
MIRNA	hsa-miR-6131	MIRNA:hsa-miR-6131	161	12028	3.25E-02	75	9	1.00E+00	88	3	1.00E+00	53	2	1.00E+00	44	4
MIRNA	hsa-miR-619-5p	MIRNA:hsa-miR-619-5p	311	12028	1.62E-01	75	11	3.19E-04	88	15	3.75E-01	53	9	1.00E+00	44	3
MIRNA	hsa-miR-6499-3p	MIRNA:hsa-miR-6499-3p	646	12028	2.93E-03	75	18	1.38E-07	88	25	1.00E+00	53	11	7.43E-01	44	11
MIRNA	hsa-miR-6506-5p	MIRNA:hsa-miR-6506-5p	286	12028	5.23E-01	75	10	1.02E-04	88	15	1.00E+00	53	8	1.00E+00	44	3

MIRNA	hsa-miR-6512-3p	MIRNA:hsa-miR-6512-3p	354	12028	8.63E-02	75	12	1.27E-02	88	14	1.00E+00	53	9	1.00E+00	44	2
MIRNA	hsa-miR-6513-5p	MIRNA:hsa-miR-6513-5p	283	12028	6.48E-02	75	11	8.78E-07	88	17	1.00E+00	53	6	1.00E+00	44	7
MIRNA	hsa-miR-6516-5p	MIRNA:hsa-miR-6516-5p	466	12028	1.00E+00	75	11	5.18E-06	88	20	1.00E+00	53	5	1.00E+00	44	8
MIRNA	hsa-miR-6720-5p	MIRNA:hsa-miR-6720-5p	355	12028	8.89E-02	75	12	1.31E-02	88	14	1.00E+00	53	9	1.00E+00	44	2
MIRNA	hsa-miR-6741-3p	MIRNA:hsa-miR-6741-3p	163	12028	3.74E-01	75	8	1.08E-04	88	12	1.00E+00	53	1	1.00E+00	-1	-1
MIRNA	hsa-miR-6778-3p	MIRNA:hsa-miR-6778-3p	476	12028	9.39E-03	75	15	9.64E-07	88	21	5.74E-05	53	15	1.00E+00	44	8
MIRNA	hsa-miR-6779-5p	MIRNA:hsa-miR-6779-5p	476	12028	5.83E-02	75	14	1.38E-09	88	24	1.00E+00	53	10	1.00E+00	44	9
MIRNA	hsa-miR-6780a-5p	MIRNA:hsa-miR-6780a-5p	472	12028	5.27E-02	75	14	1.14E-09	88	24	1.00E+00	53	10	1.00E+00	44	9
MIRNA	hsa-miR-6785-5p	MIRNA:hsa-miR-6785-5p	578	12028	5.56E-01	75	14	5.17E-06	88	22	1.00E+00	53	9	1.00E+00	44	7
MIRNA	hsa-miR-6788-5p	MIRNA:hsa-miR-6788-5p	383	12028	1.95E-01	75	12	1.48E-07	88	20	1.00E+00	53	8	1.00E+00	44	7
MIRNA	hsa-miR-6791-3p	MIRNA:hsa-miR-6791-3p	226	12028	1.00E+00	75	7	4.46E-04	88	13	1.00E+00	53	5	1.00E+00	44	3
MIRNA	hsa-miR-6799-5p	MIRNA:hsa-miR-6799-5p	483	12028	3.89E-01	75	13	1.54E-07	88	22	1.00E+00	53	10	1.00E+00	44	8
MIRNA	hsa-miR-6807-5p	MIRNA:hsa-miR-6807-5p	418	12028	1.00E+00	75	9	7.62E-11	88	24	1.00E+00	53	8	1.00E+00	44	5
MIRNA	hsa-miR-6808-5p	MIRNA:hsa-miR-6808-5p	488	12028	1.00E+00	75	11	8.42E-05	88	19	3.56E-01	53	11	1.00E+00	44	7

MIRNA	hsa-miR-6827-3p	MIRNA:hsa-miR-6827-3p	86	12028	1.00E+00	75	4	9.90E-03	88	8	1.00E+00	53	2	1.00E+00	44	1
MIRNA	hsa-miR-6829-3p	MIRNA:hsa-miR-6829-3p	226	12028	1.00E+00	75	7	4.46E-04	88	13	1.00E+00	53	5	1.00E+00	44	3
MIRNA	hsa-miR-6836-3p	MIRNA:hsa-miR-6836-3p	139	12028	1.00E+00	75	5	3.74E-02	88	9	1.00E+00	53	4	1.00E+00	44	2
MIRNA	hsa-miR-6840-3p	MIRNA:hsa-miR-6840-3p	331	12028	1.00E+00	75	8	5.55E-03	88	14	1.00E+00	53	8	1.00E+00	44	6
MIRNA	hsa-miR-6849-3p	MIRNA:hsa-miR-6849-3p	505	12028	1.00E+00	75	12	9.59E-04	88	18	7.29E-02	53	12	1.00E+00	44	4
MIRNA	hsa-miR-6883-5p	MIRNA:hsa-miR-6883-5p	578	12028	5.56E-01	75	14	5.17E-06	88	22	1.00E+00	53	9	1.00E+00	44	7
MIRNA	hsa-miR-6884-5p	MIRNA:hsa-miR-6884-5p	386	12028	1.00E+00	75	10	3.60E-02	88	14	1.00E+00	53	8	1.00E+00	44	8
MIRNA	hsa-miR-6890-3p	MIRNA:hsa-miR-6890-3p	325	12028	5.30E-04	75	14	8.03E-07	88	18	6.28E-02	53	10	1.00E+00	44	4
MIRNA	hsa-miR-6893-5p	MIRNA:hsa-miR-6893-5p	486	12028	1.00E+00	75	11	5.29E-04	88	18	3.42E-01	53	11	1.00E+00	44	7
MIRNA	hsa-miR-7106-5p	MIRNA:hsa-miR-7106-5p	543	12028	1.00E+00	75	13	7.62E-05	88	20	1.00E+00	53	9	1.00E+00	44	9
MIRNA	hsa-miR-7107-5p	MIRNA:hsa-miR-7107-5p	118	12028	3.30E-02	75	8	1.00E+00	88	4	1.00E+00	53	2	1.00E+00	44	2
MIRNA	hsa-miR-7151-3p	MIRNA:hsa-miR-7151-3p	307	12028	1.43E-01	75	11	1.63E-02	88	13	2.55E-05	53	13	1.00E+00	44	3
MIRNA	hsa-miR-766-3p	MIRNA:hsa-miR-766-3p	435	12028	1.00E+00	75	11	4.30E-03	88	16	8.34E-01	53	10	1.00E+00	44	6
MIRNA	hsa-miR-7703	MIRNA:hsa-miR-7703	331	12028	4.26E-02	75	12	1.00E+00	88	11	1.00E+00	53	7	1.00E+00	44	1

MIRNA	hsa-miR-7977	MIRNA:hsa-miR-7977	464	12028	6.74E-03	75	15	2.57E-04	88	18	3.63E-03	53	13	3.03E-02	44	11
MIRNA	hsa-miR-887-5p	MIRNA:hsa-miR-887-5p	349	12028	1.03E-02	75	13	2.62E-06	88	18	9.44E-01	53	9	1.90E-01	44	9
MIRNA	hsa-miR-940	MIRNA:hsa-miR-940	525	12028	1.00E+00	75	11	1.74E-03	88	18	7.09E-01	53	11	1.00E+00	44	7
MIRNA	MIRNA root	MIRNA:000000	9517	12028	1.00E+00	75	71	1.26E-03	88	87	1.00E+00	53	52	1.00E+00	44	43

*Table 7 Results of fitting sinusoidal curve to differential age in S- verse M-phase cells, using the equation Delta ~ A * sin(B * (Age - C)) + D.*

Gene	A	B	C	D	B_pvalue	Period	Resid_standard_error	bootstrap_p_value
BPNT1	4.08E-02	9.83E-01	5.87E+00	-5.31E-03	2.27E-04	6.40E+00	2.80E-02	2.81E-01
CEBPZOS	4.08E-02	9.85E-01	2.22E+00	3.40E-03	3.31E-04	6.40E+00	3.10E-02	8.19E-01
CWF19L1	3.87E-02	5.83E-01	-1.74E+00	-4.50E-03	1.53E-03	1.08E+01	2.60E-02	5.90E-01
DCAF16	5.77E-02	7.56E-01	2.37E+00	-2.21E-03	4.76E-03	8.30E+00	6.60E-02	9.89E-01
DFFA	2.57E-02	8.07E-01	2.08E+00	7.78E-04	5.31E-03	7.80E+00	2.80E-02	8.93E-01
GINS1	1.61E-02	7.61E-01	6.55E+00	4.98E-04	1.29E-02	8.30E+00	2.50E-02	7.38E-01
GINS4	5.17E-02	7.11E-01	7.27E+00	2.76E-03	1.87E-03	8.80E+00	4.00E-02	7.53E-01
GNL3L	-3.59E-02	5.89E-01	-1.20E+00	6.15E-03	3.80E-02	1.07E+01	6.20E-02	9.67E-01
HELLS	1.92E-02	5.07E-01	-3.01E+00	-2.52E-03	8.78E-04	1.24E+01	1.00E-02	4.40E-02
INIP	6.62E-02	8.34E-01	4.35E-01	-1.22E-04	9.47E-06	7.50E+00	2.00E-02	9.60E-02
IPP	-2.59E-02	7.73E-01	4.44E-01	2.06E-04	4.59E-03	8.10E+00	3.00E-02	2.89E-01
LYRM7	4.28E-02	6.96E-01	7.67E-01	-5.17E-03	2.88E-04	9.00E+00	2.50E-02	4.94E-01
MDM2	-2.23E-02	4.25E-01	-6.52E+00	2.04E-04	4.53E-04	1.48E+01	9.00E-03	4.70E-02
MRI1	5.68E-02	1.05E+00	3.93E+00	6.23E-03	1.41E-04	6.00E+00	4.20E-02	6.63E-01
NUP43	2.74E-02	7.57E-01	-4.92E-01	1.61E-04	1.93E-02	8.30E+00	4.50E-02	8.49E-01
PAICS	1.82E-02	7.37E-01	1.92E+00	-1.24E-03	3.77E-03	8.50E+00	2.00E-02	8.37E-01
PHAX	2.72E-02	7.16E-01	1.64E+00	-2.71E-03	4.18E-04	8.80E+00	1.70E-02	4.84E-01

RBBP4	2.19E-02	6.21E-01	-1.64E+00	-1.58E-03	5.39E-03	1.01E+01	2.00E-02	2.91E-01
RBM8A	2.54E-02	7.79E-01	1.90E+00	-2.46E-04	5.02E-03	8.10E+00	2.90E-02	3.81E-01
RPL7L1	5.97E-03	9.03E-01	1.20E+00	1.07E-04	1.62E-04	7.00E+00	4.00E-03	5.09E-01
SNRPD3	-3.02E-02	5.69E-01	-1.87E+00	4.26E-03	9.75E-03	1.10E+01	3.30E-02	3.44E-01
SPC24	2.31E-02	7.41E-01	-4.98E+00	-2.78E-04	6.29E-04	8.50E+00	1.40E-02	6.00E-02
TOP3A	2.06E-02	7.31E-01	3.11E+00	-6.87E-04	9.47E-04	8.60E+00	1.40E-02	4.87E-01
XRCC2	1.82E-02	5.97E-01	-1.32E+00	-2.65E-03	8.62E-03	1.05E+01	2.00E-02	3.86E-01
ZNF714	-2.03E-02	6.99E-01	-6.16E-01	6.36E-04	8.83E-03	9.00E+00	2.20E-02	9.17E-01

Table 8 All plasmids used in this work

ID	Name	Description
pAG018	pCMV-rtTA3	rtTA3 transactivator expressed downstream of the CMV promoter.
pAG051	pTRE2-ATG14-IRES-GFP	Bicitronic mammalian expression vector with TET-ON promoter. Tetracycline inducible promoter upstream of ATG14 (NM_014924.5)-IRES-GFP. Beta-globin intron and poly-A signal follows GFP. Ampicillin resistance selection.
pAG053	pTRE2-BVES-IRES-GFP	Bicitronic mammalian expression vector with TET-ON promoter. Tetracycline inducible promoter upstream of BVES (NM_001199563.2)-IRES-GFP. Beta-globin intron and poly-A signal follows GFP. Ampicillin resistance selection.
pAG054	pTRE2-ACBD7-IRES-GFP	Bicitronic mammalian expression vector with TET-ON promoter. Tetracycline inducible promoter upstream of ACBD7 (NM_001039844)-IRES-GFP. Beta-globin intron and poly-A signal follows GFP. Ampicillin resistance selection.
pAG045	CAG::mScarlet-P2A-TadA8.20	TadA8.20 from Xiao et al. 2023 was cloned into a pAAV-CAG-mScarlet-P2A vector
pAG047	CAG::mScarlet-p2a-TadA8.20-32aaGGS-2Nlamdba	TadA8.20 from Xiao et al. 2023 was cloned into a pAAV-CAG-mScarlet-P2A vector with Nlamdba fused to its C-terminus via a 32 amino acid GGS linker.
pAG048	CAG::mScarlet-p2a-TadA7.10	TadA7.10 (ABE7.10) from Gaudelli et al. 2020 was cloned into a pAAV-CAG-mScarlet-P2A vector.
pAG049	CAG::mScarlet-p2a-TadA8.20-32aaGGS-DpsN	TadA8.20 from Xiao et al. 2023 was cloned into a pAAV-CAG-mScarlet-P2A vector with E.coli peptide Dps-N fused to its N-terminus via a 6 amino acid GGS linker.
pAG001	BE3-P2A-EGFP	BE3 from Grunewald et al. 2019
pAG002	pCMV_ABEmax_P2A_GFP	ABEmax from Gaudelli et al. 2017
pAG003	ADAR2(E488Q)-NLambda	hADAR2_E488Q_(CD)—λN + BoxB gRNA ASO (from Montiel-González et al.) Courtesy of Rory Maizels
pAG004	Cas13-ADAR2(E488Q)	hADAR2_E488Q_(CD)—pspCas13b + Cas1b_gRNA ASO (from Cox et al.). Courtesy of Rory Maizels
pAG005	PABP-ADAR2(E488Q)	hADAR2_E488Q_(CD)—PABP_(RRM***) + polyU ASO - ***PolyA Binding Protein's RNA recognition motif only. Courtesy of Rory Maizels
pAG006	SNAP-ADAR2(E488Q)	hADAR2_SNAP + BG-gRNA ASO (from Vogel et al.) Courtesy of Rory Maizels

Table 9 All primers, G-blocks and Ultramers used in this study.

ID	Description	Sequence
AG275	amplfiy ATG14 from HEK cDNA	GCTGACGCGTGCTAGAGGCCATCATGGCG
AG276	amplfiy ATG14 from HEK cDNA	TATCGATACCGTCGATACTGTTAAAAGATTTATTGC
AG387	repair ATG14 editing sites	ACAGTATCGACGGTATCGATAAGCTAATTCCGCC
AG388	repair ATG14 editing sites	TGGCCTCTAGCACGCGTCAGCTGACTAGAGG
AG415	AmpR casette	AAGTAAGTTGGCCGCAGTGTTATCACTC
AG416	AmpR casette	ACACTGCGGCCAACTTACTTCTGACAAC
AG421	repair ATG14 editing sites	AAGCAATTCTTCCATCTCAACCTCCCGAGTAGCTGGGATTATAGGCACC
AG422	repair ATG14 editing sites	TAGCACTTTAGGAGGCCAAGGCAAGCGGATCACCTGAAGTCAGGAGTTCAAGAC
AG423	repair ATG14 editing sites	GTTGAGATGGAAGAATTGCTTGAACCCAGGAGATGGTGGTTGCAGTG
AG424	repair ATG14 editing sites	GCGGCAGGAGAATCGCTTGAACCGGGAGGCAGAGGTTGCAG
AG425	repair ATG14 editing sites	TCAAGCGATTCTCCTGCCGCAGCCTCCCGAGTAGCTGGG
AG426	repair ATG14 editing sites	CTTGGCCTCCTAAAGTGCTAGGATTACAGGCCTGAGCCACCGCG
AG353	amplfiy BVES from HEK cDNA	GCTTATCGATACCGTCGACGGCTCTATCCAGACAAAAAGCATCT
AG408	repair BVES editing sites	GCCCGACTAGCACGCGTCAGCTGACTAGAGG
AG427	repair BVES editing sites	CACATTCTCAGCTCACTGCAAACTCCACCTCCTGGATTCACGCC
AG428	repair BVES editing sites	CGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTAC
AG429	repair BVES editing sites	ACTGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCC
AG430	repair BVES editing sites	TCCATGGCCTGGCCTGTAGTAAACCCTCTGCCTTCCACTGTGTTCACGTTCGC
AG431	repair BVES editing sites	TACTACAGGCCAGGCCATGGACACCCCACCCAGGTACTGCAGTGAGGG
AG432	repair BVES editing sites	AGAGATGGGGTTTAGCCATGTTGCCCAGGCTGGTCTTGAACTCCTGGGCTTAAGCAATCCACCCACCTTGGCCCCCC
AG433	repair BVES editing sites	CATGGCTAAACCCCATCTCTACAAAAAAAAATTAGCTGGGTGTGGTGGTACATACCTGTAGTCCCAGCTACTCAGAAGGCTG
AG434	repair BVES editing sites	TTGCAGTGAGCTGAGAATGTGCCACTACACTCCAGCCTGGGCGACAGAG
AG352	repair BVES editing sites	TGACGCGTGCTAGTCGGGCTCCCGTCGAG
AG385	amplfiy BVES from HEK cDNA	TCGACGGTATCGATAAGCTAATTCCGCC
AG266	amplify ACBD7 from HEK cDNA	TATCGATACCGTCGATAGCCTTTTACCCAATTTCC
AG269	amplify ACBD7 from HEK cDNA	GCTGACGCGTGCTAGGAACTGAAAGAACTCTATGG

AG385	repair ACBD7 editing sites	TCGACGGTATCGATAAGCTAATTCCGCC
AG386	repair ACBD7 editing sites	CTAGCACGCGTCAGCTGACTAGAGG
AG436	repair ACBD7 editing sites	CTTCTGACCTCAGTAATCCACCTACCTCGGCCTCCCAAAGTGCCGAGATTGGAGG
AG437	repair ACBD7 editing sites	GGTTGCAGTGAGCCAAGATCACACCACTGCACTCCAGCCTGGGCGAC
AG438	repair ACBD7 editing sites	GGGTACAGCAAACCACCATGGCACATGTATACCTATGTAAC
AG439	repair ACBD7 editing sites	CCATGGTGGTTTGCTGTACCCATCAACCGGTCACCTACATCAGG
AG456	repair ACBD7 editing sites	GTAGGTGGATTACTGAGGTCAGAAGTTCGAGACC
AG457	repair ACBD7 editing sites	GGTGTGATCTTGGCTCACTGCAACCTCC
AG435	repair ACBD7 editing sites - ultramer	AGGTGGATTACTGAGGTCAGAAGTTCGAGACCAGCCTGACCAACATTGCAAAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGCCATTGGTGGTGCACACCTGTAATTCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTTTACCCAGGAGATGGAGGTTGCAGTGAGCCAAGATCACA
ONT_RT002	original Oxford Nanopore poly-T RT primer	ACTTGCCTGTCGCTCTATCTTCTTTTTTTTTTTTTTTTTTVN
ONT_SSPII002	original Oxford Nanopore TSO from cDNA PCR kit with UMI	TTTCTGTTGGTGCTGATATTGCNNNNNNNNTTTVVVVTTVVVVTTVVVVTTVVVVTTTmGmGmG
ONT_neb_BC_001	SMART-seq barcoding primer	/5Phos/GGT GCT GAA GAA AGT TGT CGG TGT CTT TGT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_002	SMART-seq barcoding primer	/5Phos/GGT GCT GTC GAT TCC GTT TGT AGT CGT CTG TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_003	SMART-seq barcoding primer	/5Phos/GGT GCT GGA GTC TTG TGT CCC AGT TAC CAG GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_004	SMART-seq barcoding primer	/5Phos/GGT GCT GTT CGG ATT CTA TCG TGT TTC CCT ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_005	SMART-seq barcoding primer	/5Phos/GGT GCT GCT TGT CCA GGG TTT GTG TAA CCT TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_006	SMART-seq barcoding primer	/5Phos/GGT GCT GTT CTC GCA AAG GCA GAA AGT AGT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_007	SMART-seq barcoding primer	/5Phos/GGT GCT GGT GTT ACC GTG GGA ATG AAT CCT TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_008	SMART-seq barcoding primer	/5Phos/GGT GCT GTT CAG GGA ACA AAC CAA GTT ACG TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_009	SMART-seq barcoding primer	/5Phos/GGT GCT GAA CTA GGC ACA GCG AGT CTT GGT TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_010	SMART-seq barcoding primer	/5Phos/GGT GCT GAA GCG TTG AAA CCT TTG TCC TCT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_011	SMART-seq barcoding primer	/5Phos/GGT GCT GGT TTC ATC TAT CGG AGG GAA TGG ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_012	SMART-seq barcoding primer	/5Phos/GGT GCT GCA GGT AGA AAG AAG CAG AAT CGG ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_013	SMART-seq barcoding primer	/5Phos/GGT GCT GAG AAC GAC TTC CAT ACT CGT GTG ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_014	SMART-seq barcoding primer	/5Phos/GGT GCT GAA CGA GTC TCT TGG GAC CCA TAG ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T

ONT_neb_BC_0 15	SMART-seq barcoding primer	/5Phos/GGT GCT GAG GTC TAC CTC GCT AAC ACC ACT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 16	SMART-seq barcoding primer	/5Phos/GGT GCT GCG TCA ACT GAC AGT GGT TCG TAC TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 17	SMART-seq barcoding primer	/5Phos/GGT GCT GAC CCT CCA GGA AAG TAC CTC TGA TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 18	SMART-seq barcoding primer	/5Phos/GGT GCT GCC AAA CCC AAC AAC CTA GAT AGG CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 19	SMART-seq barcoding primer	/5Phos/GGT GCT GGT TCC TCG TGC AGT GTC AAG AGA TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 20	SMART-seq barcoding primer	/5Phos/GGT GCT GTT GCG TCC TGT TAC GAG AAC TCA TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 21	SMART-seq barcoding primer	/5Phos/GGT GCT GGA GCC TCT CAT TGT CCG TTC TCT ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 22	SMART-seq barcoding primer	/5Phos/GGT GCT GAC CAC TGC CAT GTA TCA AAG TAC GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 23	SMART-seq barcoding primer	/5Phos/GGT GCT GCT TAC TAC CCA GTG AAC CTC CTC GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 24	SMART-seq barcoding primer	/5Phos/GGT GCT GGC ATA GTT CTG CAT GAT GGG TTA GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 25	SMART-seq barcoding primer	/5Phos/GGT GCT GGT AAG TTG GGT ATG CAA CGC AAT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 26	SMART-seq barcoding primer	/5Phos/GGT GCT GCA TAC AGC GAC TAC GCA TTC TCA TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 27	SMART-seq barcoding primer	/5Phos/GGT GCT GCG ACG GTT AGA TTC ACC TCT TAC ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 28	SMART-seq barcoding primer	/5Phos/GGT GCT GTG AAA CCT AAG AAG GCA CCG TAT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 29	SMART-seq barcoding primer	/5Phos/GGT GCT GCT AGA CAC CTT GGG TTG ACA GAC CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 30	SMART-seq barcoding primer	/5Phos/GGT GCT GTC AGT GAG GAT CTA CTT CGA CCC ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 31	SMART-seq barcoding primer	/5Phos/GGT GCT GTG CGT ACA GCA ATC AGT TAC ATT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 32	SMART-seq barcoding primer	/5Phos/GGT GCT GCC AGT AGA AGT CCG ACA ACG TCA TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 33	SMART-seq barcoding primer	/5Phos/GGT GCT GCA GAC TTG GTA CGG TTG GGT AAC TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 34	SMART-seq barcoding primer	/5Phos/GGT GCT GGG ACG AAG AAC TCA AGT CAA AGG CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 35	SMART-seq barcoding primer	/5Phos/GGT GCT GCT ACT TAC GAA GCT GAG GGA CTG CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 36	SMART-seq barcoding primer	/5Phos/GGT GCT GAT GTC CCA GTT AGA GGA GGA AAC ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 37	SMART-seq barcoding primer	/5Phos/GGT GCT GGC TTG CGA TTG ATG CTT AGT ATC ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 38	SMART-seq barcoding primer	/5Phos/GGT GCT GAC CAC AGG AGG ACG ATA CAG AGA ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 39	SMART-seq barcoding primer	/5Phos/GGT GCT GCC ACA GTG TCA ACT AGA GCC TCT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 40	SMART-seq barcoding primer	/5Phos/GGT GCT GTA GTT TGG ATG ACC AAG GAT AGC CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 41	SMART-seq barcoding primer	/5Phos/GGT GCT GGG AGT TCG TCC AGA GAA GTA CAC GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T

ONT_neb_BC_042	SMART-seq barcoding primer	/5Phos/GGT GCT GCT ACG TGT AAG GCA TAC CTG CCA GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_043	SMART-seq barcoding primer	/5Phos/GGT GCT GCT TTC GTT GTT GAC TCG ACG GTA GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_044	SMART-seq barcoding primer	/5Phos/GGT GCT GAG TAG AAA GGG TTC CTT CCC ACT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_045	SMART-seq barcoding primer	/5Phos/GGT GCT GGA TCC AAC AGA GAT GCC TTC AGT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_046	SMART-seq barcoding primer	/5Phos/GGT GCT GGC TGT GTT CCA CTT CAT TCT CCT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_047	SMART-seq barcoding primer	/5Phos/GGT GCT GGT GCA ACT TTC CCA CAG GTA GTT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_048	SMART-seq barcoding primer	/5Phos/GGT GCT GCA TCT GGA ACG TGG TAC ACC TGT ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_049	SMART-seq barcoding primer	/5Phos/GGT GCT GAC TGG TGC AGC TTT GAA CAT CTA GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_050	SMART-seq barcoding primer	/5Phos/GGT GCT GAT GGA CTT TGG TAA CTT CCT GCG TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_051	SMART-seq barcoding primer	/5Phos/GGT GCT GGT TGA ATG AGC CTA CTG GGT CCT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_052	SMART-seq barcoding primer	/5Phos/GGT GCT GTG AGA GAC AAG ATT GTT CGT GGA CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_053	SMART-seq barcoding primer	/5Phos/GGT GCT GAG ATT CAG ACC GTC TCA TGC AAA GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_054	SMART-seq barcoding primer	/5Phos/GGT GCT GCA AGA GCT TTG ACT AAG GAG CAT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_055	SMART-seq barcoding primer	/5Phos/GGT GCT GTG GAA GAT GAG ACC CTG ATC TAC GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_056	SMART-seq barcoding primer	/5Phos/GGT GCT GTC ACT ACT CAA CAG GTG GCA TGA ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_057	SMART-seq barcoding primer	/5Phos/GGT GCT GGC TAG GTC AAT CTC CTT CGG AAG TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_058	SMART-seq barcoding primer	/5Phos/GGT GCT GCA GGT TAC TCC TCC GTG AGT CTG ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_059	SMART-seq barcoding primer	/5Phos/GGT GCT GTC AAT CAA GAA GGG AAA GCA AGG TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_060	SMART-seq barcoding primer	/5Phos/GGT GCT GCA TGT TCA ACC AAG GCT TCT ATG GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_061	SMART-seq barcoding primer	/5Phos/GGT GCT GAG AGG GTA CTA TGT GCC TCA GCA CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_062	SMART-seq barcoding primer	/5Phos/GGT GCT GCA CCC ACA CTT ACT TCA GGA CGT ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_063	SMART-seq barcoding primer	/5Phos/GGT GCT GTT CTG AAG TTC CTG GGT CTT GAA CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_064	SMART-seq barcoding primer	/5Phos/GGT GCT GGA CAG ACA CCG TTC ATC GAC TTT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_065	SMART-seq barcoding primer	/5Phos/GGT GCT GTT CTC AGT CTT CCT CCA GAC AAG GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_066	SMART-seq barcoding primer	/5Phos/GGT GCT GCC GAT CCT TGT GGC TTC TAA CTT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_067	SMART-seq barcoding primer	/5Phos/GGT GCT GGT TTG TCA TAC TCG TGT GCT CAC CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_068	SMART-seq barcoding primer	/5Phos/GGT GCT GGA ATC TAA GCA AAC ACG AAG GTG GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T

ONT_neb_BC_0 69	SMART-seq barcoding primer	/5Phos/GGT GCT GTA CAG TCC GAG CCT CAT GTG ATC TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 70	SMART-seq barcoding primer	/5Phos/GGT GCT GAC CGA GAT CCT ACG AAT GGA GTG TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 71	SMART-seq barcoding primer	/5Phos/GGT GCT GCC TGG GAG CAT CAG GTA GTA ACA GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 72	SMART-seq barcoding primer	/5Phos/GGT GCT GTA GCT GAC TGT CTT CCA TAC CGA CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 73	SMART-seq barcoding primer	/5Phos/GGT GCT GAA GAA ACA GGA TGA CAG AAC CCT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 74	SMART-seq barcoding primer	/5Phos/GGT GCT GTA CAA GCA TCC CAA CAC TTC CAC TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 75	SMART-seq barcoding primer	/5Phos/GGT GCT GGA CCA TTG TGA TGA ACC CTG TTG TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 76	SMART-seq barcoding primer	/5Phos/GGT GCT GAT GCT TGT TAC ATC AAC CCT GGA CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 77	SMART-seq barcoding primer	/5Phos/GGT GCT GCG ACC TGT TTC TCA GGG ATA CAA CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 78	SMART-seq barcoding primer	/5Phos/GGT GCT GAA CAA CCG AAC CTT TGA ATC AGA ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 79	SMART-seq barcoding primer	/5Phos/GGT GCT GTC TCG GAG ATA GTT CTC ACT GCT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 80	SMART-seq barcoding primer	/5Phos/GGT GCT GCG GAT GAA CAT AGG ATA GCG ATT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 81	SMART-seq barcoding primer	/5Phos/GGT GCT GCC TCA TCT TGT GAA GTT GTT TCG GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 82	SMART-seq barcoding primer	/5Phos/GGT GCT GAC GGT ATG TCG AGT TCC AGG ACT ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 83	SMART-seq barcoding primer	/5Phos/GGT GCT GTG GCT TGA TCT AGG TAA GGT CGA ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 84	SMART-seq barcoding primer	/5Phos/GGT GCT GGT AGT GGA CCT AGA ACC TGT GCC ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 85	SMART-seq barcoding primer	/5Phos/GGT GCT GAA CGG AGG AGT TAG TTG GAT GAT CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 86	SMART-seq barcoding primer	/5Phos/GGT GCT GAG GTG ATC CCA ACA AGC GTA AGT ATT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 87	SMART-seq barcoding primer	/5Phos/GGT GCT GTA CAT GCT CCT GTT GTT AGG GAG GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 88	SMART-seq barcoding primer	/5Phos/GGT GCT GTC TTC TAC TAC CGA TCC GAA GCA GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 89	SMART-seq barcoding primer	/5Phos/GGT GCT GAC AGC ATC AAT GTT TGG CTA GTT GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 90	SMART-seq barcoding primer	/5Phos/GGT GCT GGA TGT AGA GGG TAC GGT TTG AGG CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 91	SMART-seq barcoding primer	/5Phos/GGT GCT GGG CTC CAT AGG AAC TCA CGC TAC TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 92	SMART-seq barcoding primer	/5Phos/GGT GCT GTT GTG AGT GGA AAG ATA CAG GAC CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 93	SMART-seq barcoding primer	/5Phos/GGT GCT GAG TTT CCA TCA CTT CAG ACT TGG GTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 94	SMART-seq barcoding primer	/5Phos/GGT GCT GGA TTG TCC TCA AAC TGC CAC CTA CTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
ONT_neb_BC_0 95	SMART-seq barcoding primer	/5Phos/GGT GCT GCC TGT CTG GAA GAA GAA TGG ACT TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T

ONT_neb_BC_096	SMART-seq barcoding primer	/5Phos/GGT GCT GCT GAA CGG TCA TAG AGT CCA CCA TTT AAC CTA AGC AGT GGT ATC AAC GCA GAG T
AG360	cloning gen2 editors	CGACTTCGGAGTTTGCAGCTTTCCATTG
AG361	cloning gen2 editors	GAAAGCTGCAAACTCCGAAGTCGAGTTTTCC
AG362	cloning gen2 editors	CGATAAGCTTGATATCGTCAGTAGAGGATTGTGC
AG363	cloning gen2 editors	CCCTGGACCTTCCGAAGTCGAGTTTTCC
AG364	cloning gen2 editors	CGACTTCGGAAGGTCCAGGGTTCTCCTC
AG365	cloning gen2 editors	CGACTTCGGAAGGTCCAGGGTTCTCCTCC
AG366	cloning gen2 editors	GAATTGGTGCTAGCGCCACCATGGTGAGCAAG
AG367	cloning gen2 editors	ACCCTGGACCTTCCGAAGTCGAGTTTTCCC
AG368	cloning gen2 editors	TACGGGCATTAGAACCACCAGAAGAACCACC
AG369	cloning gen2 editors	TGGTGGTTCTAATGCCCGTACGCG
AG370	cloning gen2 editors	ATAAGCTTGATATCGTTTGCAGCTTTCCATTGAG
AG371	cloning gen2 editors	CGACTTCGGACATAGGTCCAGGGTTCTC
AG372	cloning gen2 editors	CCCTGGACCTATGTCCGAAGTCGAGTTTTC
AG373	cloning gen2 editors	CGATAAGCTTGATATCAGAACCACCAGAAGAAC
AG379	cloning gen2 editors	AGGTCCAGGGTTCTCCTCCAC
AG380	cloning gen2 editors	GTCTCATCATTTTGGCAAAGAATTGGTG
AG381	Gblock of TadA8.20-32aaGGS-4Nlambda (codon optimised)	GCTGGAGACGTGGAGGAGAACCCTGGACCTTCCGAAGTCGAATTTTCTCACGAATACTGGATGCGGCATGCGTTGACCCTGGCGAAGCGCGC CCGGGACGAACGGGAGGTCCCTGTAGGCGCGGTGCTTGTTCTTAATAATAGGGTCATCGGCGAAGGTTGGAATCGCGCAATTGGTCTCCATG ACCCCACCGCACATGCCGAAATCATGGCTTTGCGCCAAGGAGGTTTGGTCATGCAGAACTATAGGCTGTATGATGCAACTCTGTATAGTACT TTTGAGCCTTGTGTAATGTGTGCTGGAGCTATGATCCATTCTCGAATAGGGAGAGTAGTGTTTGGGGTGCGAAATGCCAAGACCGGTGCAGC AGGATCCTTGATGGACGTTCTTCATCACCCTGGAATGAATCATAGAGTCGAGATAACTGAAGGCATACTTGCGGATGAGTGTGCAGCGCTCC TTTGTCGATTTTTTCGAATGCCTCGCAGGGTATTCAATGCGCAGAAAAAGGCCCAATCAAGCACAGACTCTGGAGGGTCTTCAGGCGGTAGC AGCGGCTCAGAAACGCCGGGTACCAGCGAAAGTGCCACACCCGAAAGTAGTGGGGGTAGCTCTGGTGGCTCCAACGCTAGGACAAGGCGGCG CGAGCGCAGGGCAGAAAAGCAGGCTCAGTGGAAAGCAGCGAACGGAGGTGGAGGGTCAGGGGGTGGAGGGTCTGGCGGTGGAGGCTCTAATG CACGAACTAGGCGAAGGGAGCGGAGAGCAGAAAAGCAAGCGCAATGGAAGGCCGCCAATGGTGGTGGGGGTAGTGGTGGCGGTGGGAGTGGT GGCGGTGGATCAAACGCGCGCACAAGACGACGAGAGAGACGAGCTGAGAAACAAGCACAATGGAAGGCGGCAAATGGCGGGGGAGGCTCCGG CGGTGGCGGGTCAGGGGGAGGAGGATCTAATGCGCGAACTCGACGGCGGGAGAGGAGGGCTGAAAAACAGGCGCAATGGAAAGCGGCGAACT AAGATATCAAGCTTATCGATAATCAACCTCTG
AG382	Gblock of TadA8.20-32aaGGS-2Nlambda (codon optimised)	GCTGGAGACGTGGAGGAGAACCCTGGACCTTCCGAAGTCGAATTTTCTCACGAATACTGGATGCGGCATGCGTTGACCCTGGCGAAGCGCGC CCGGGACGAACGGGAGGTCCCTGTAGGCGCGGTGCTTGTTCTTAATAATAGGGTCATCGGCGAAGGTTGGAATCGCGCAATTGGTCTCCATG ACCCCACCGCACATGCCGAAATCATGGCTTTGCGCCAAGGAGGTTTGGTCATGCAGAACTATAGGCTGTATGATGCAACTCTGTATAGTACT TTTGAGCCTTGTGTAATGTGTGCTGGAGCTATGATCCATTCTCGAATAGGGAGAGTAGTGTTTGGGGTGCGAAATGCCAAGACCGGTGCAGC AGGATCCTTGATGGACGTTCTTCATCACCCTGGAATGAATCATAGAGTCGAGATAACTGAAGGCATACTTGCGGATGAGTGTGCAGCGCTCC TTTGTCGATTTTTTCGAATGCCTCGCAGGGTATTCAATGCGCAGAAAAAGGCCCAATCAAGCACAGACTCTGGAGGGTCTTCAGGCGGTAGC AGCGGCTCAGAAACGCCGGGTACCAGCGAAAGTGCCACACCCGAAAGTAGTGGGGGTAGCTCTGGTGGCTCCAACGCTAGGACAAGGCGGCG CGAGCGCAGGGCAGAAAAGCAGGCTCAGTGGAAAGCAGCGAACGGAGGTGGAGGGTCAGGGGGTGGAGGGTCTGGCGGTGGAGGCTCTAATG CACGAACTAGGCGAAGGGAGCGGAGAGCAGAAAAGCAAGCGCAATGGAAGGCCGCCAATTAAGATATCAAGCTTATCGATAATCAACCTCTG

AG383	Gblock of TadA8.20 (codon optimised)	GCTGGAGACGTGGAGGAGAACCCTGGACCTTCCGAAGTCGAATTTTCTCACGAATACTGGATGCGGCATGCGTTGACCCTGGCGAAGCGCGC CCGGGACGAACGGGAGGTCCCTGTAGGCGCGGTGCTTGTTCTTAATAATAGGGTCATCGGCGAAGGTTGGAATCGCGCAATTGGTCTCCATG ACCCCACCGCACATGCCGAAATCATGGCTTTGCGCCAAGGAGGTTTGGTCATGCAGAACTATAGGCTGTATGATGCAACTCTGTATAGTACT TTTGAGCCTTGTGTAATGTGTGCTGGAGCTATGATCCATTCTCGAATAGGGAGAGTAGTGTTTGGGGTGCGAAATGCCAAGACCGGTGCAGC AGGATCCTTGATGGACGTTCTTCATCACCCTGGAATGAATCATAGAGTCGAGATAACTGAAGGCATACTTGCGGATGAGTGTGCAGCGCTCC TTTGTCGATTTTTTCGAATGCCTCGCAGGGTATTCAATGCGCAGAAAAAGGCCCAATCAAGCACAGACTAAGATATCAAGCTTATCGATAAT CAACCT
AG384	Gblock of TadA7.10 (codon optimised)	ATGTCTGAGGTTGAATTCTCACATGAGTACTGGATGCGACATGCATTGACGCTTGCGAAGAGAGCTTGGGATGAACGCGAAGTACCTGTTGG CGCCGTATTGGTGCATAATAACCGGGTGATTGGGGAAGGATGGAACCGCCCCATTGGAAGGCATGATCCTACAGCGCACGCTGAAATCATGG CGTTGCGCCAGGGCGGTCTCGTTATGCAAAATTACCGGCTGATTGATGCCACCTTGTATGTTACACTTGAGCCCTGTGTTATGTGCGCAGGA GCTATGATTCACAGCAGAATAGGGAGAGTCGTTTTTGGGGCACGGGACGCTAAAACAGGCGCTGCTGGAAGTTTGATGGACGTGTTGCACCA TCCTGGAATGAACCATAGAGTTGAGATTACCGAGGGAATTCTGGCGGACGAATGTGCAGCTTTGTTGTCTGATTTTTTCCGGATGCGGCGGC AAGAGATAAAGGCCCAGAAAAAGGCACAATCTTCCACGGATAGCGGTGGTTCATCAGGAGGTTCTAGTGGTTCTGAAACGCCAGGAACGTCT GAAAGTGCGACTCCCGAGAGCAGCGGGGGATCTTCCGGGGGTTCCTCAGAGGTAGAGTTCTCACATGAGTATTGGATGAGACACGCCCTGAC ATTGGCAAAGCGCGCTAGAGACGAACGGGAAGTTCCCGTAGGAGCTGTCCTGGTGCTCAACAATAGGGTTATAGGTGAAGGCTGGAACCGGG CGATCGGGCTGCATGATCCCACCGCGCATGCCGAGATTATGGCCCTTCGGCAGGGTGGTCTTGTAATGCAAAATTACCGATTGATAGATGCG ACCCTTTACGTAACTTTCGAACCTTGTGTGATGTGCGCAGGAGCAATGATCCATAGTAGGATAGGGAGGGTCGTGTTCGGTGTTCGGAATGC TAAGACAGGTGCCGCCGGGTCATTGATGGATGTGCTGCATTACCCAGGCATGAATCATCGCGTGGAAATCACCGAGGGGATTTTGGCTGACG AGTGCGCGGCATTGCTTTGTTACTTCTTCCGGATGCCTCGGCAAGTCTTTAACGCTCAAAAAAAAGCGCAGTCCTCTACCGATTCTGGGGGT AGTAGCGGCGGTAGCTCCGGTTCTGAAACACCGGGGACAAGCGAGTCCGCGACACCCGAATCCTCCGGGGGGTCTTCCGGTGGAAGTTAA