# *Introduction to RNA editing*

## *Ernesto Picardi*

**University of Bari**
**IBIOM-CNR**

**ernesto.picardi@uniba.it**

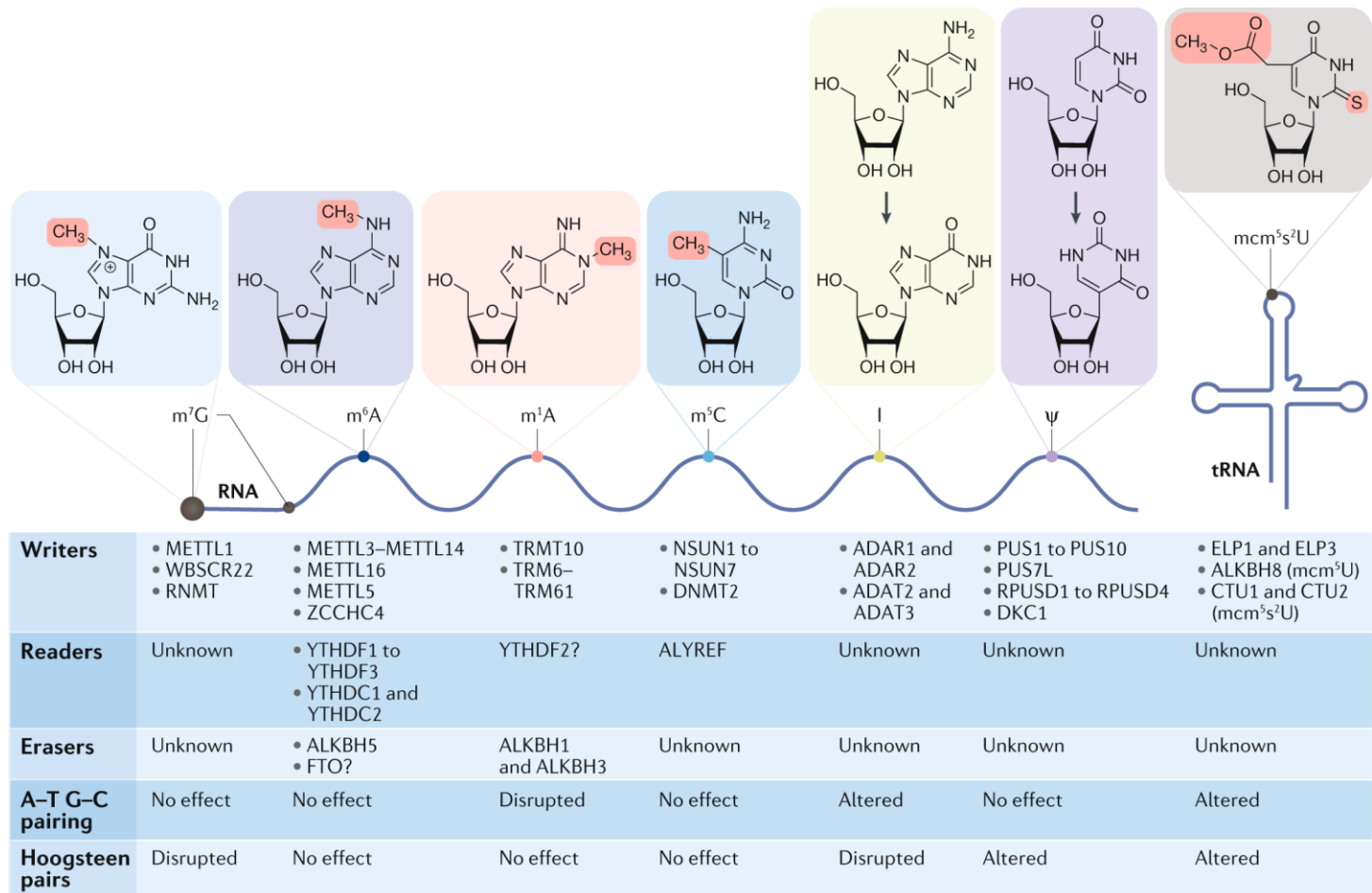www.uniba.it          www.ibiom.cnr.it

TRAINING COURSE IN
**Computational Methods**
**for Epitranscriptomics**

Bari, 11th-13th September 2024

# RNA modifications

Epitranscriptome modifications are emerging as important factors to fine tune gene expression and regulation in a variety of organisms and experimental conditions. To date more than 170 distinct chemical modifications to RNA have been characterized, including transient (i.e. m6A or m1A or m5C) and not-transient (i.e. A-to-I or C-to-U) nucleotide variants.

| | m$^7$G | m$^6$A | m$^1$A | m$^5$C | I | Ψ | |
|---|---|---|---|---|---|---|---|
| **Writers** | • METTL1<br>• WBSCR22<br>• RNMT | • METTL3–METTL14<br>• METTL16<br>• METTL5<br>• ZCCHC4 | • TRMT10<br>• TRM6–TRM61 | • NSUN1 to NSUN7<br>• DNMT2 | • ADAR1 and ADAR2<br>• ADAT2 and ADAT3 | • PUS1 to PUS10<br>• PUS7L<br>• RPUSD1 to RPUSD4<br>• DKC1 | • ELP1 and ELP3<br>• ALKBH8 (mcm$^5$U)<br>• CTU1 and CTU2 (mcm$^5$s$^2$U) |
| **Readers** | Unknown | • YTHDF1 to YTHDF3<br>• YTHDC1 and YTHDC2 | YTHDF2? | ALYREF | Unknown | Unknown | Unknown |
| **Erasers** | Unknown | • ALKBH5<br>• FTO? | ALKBH1 and ALKBH3 | Unknown | Unknown | Unknown | Unknown |
| **A–T G–C pairing** | No effect | No effect | Disrupted | No effect | Altered | No effect | Altered |
| **Hoogsteen pairs** | Disrupted | No effect | No effect | No effect | Disrupted | Altered | Altered |

Barbieri and Kouzarides 2020 Nat. Rev. Cancer

# RNA editing

RNA editing is a widespread co- and post-transcriptional molecular phenomenon that can increase the complexity of the eukaryotic transcriptome and proteome through a variety of mechanistically and evolutionarily unrelated pathways.
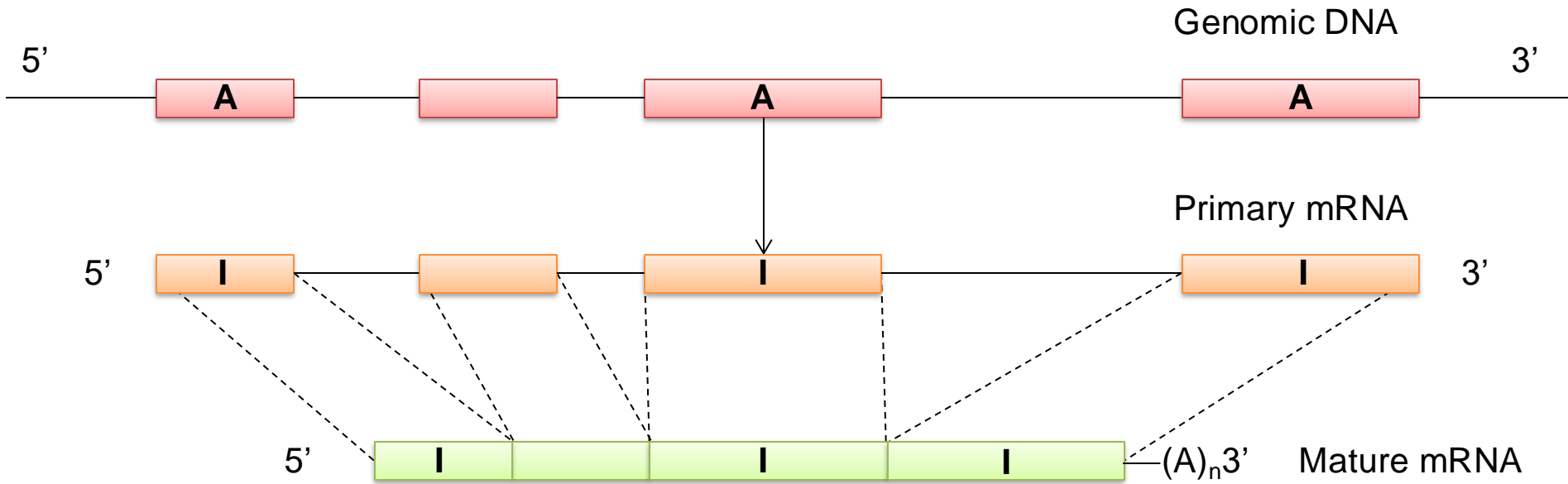


Primary RNAs are modified at specific positions by <u>base substitutions</u> or <u>insertions and deletions</u>.

# Insertion/Deletion RNA editing

## Alignment for EDI_000000649

```
Genomic:  ttacgcctattgatcatactatatgtata--a---------a-acg-gg-----attta-a
          ||||||||||||||||||||||||||||||    |    |   ||  ||     ||||| |
cDNA:     ttacgcctattgatcatactatatgtataTTaTTTTTTTTTTaTacgTggTTTTTatttaTa
Protein:   Y  A  Y  W  S  Y  Y  M  Y  I  I  F  F  I  R  G  F  Y  L  Y

Genomic:  -g-g---a-gtttac-TT-g---a---a------g-g-g---a--a-a-aagtcctcga--
          | |    | ||||||  |    |   |     | | |    |  | |  ||||||||||  |
cDNA:     Tg-gTTTaTgtttac--gTTTaTaTaTTTTTgTgTgTTTaTTaTaTaTaagtcctcgaTTT
Protein:   G  L  C  L     R  L  F  I  F  L  C  V  Y  Y  I  S  P  R  L

Genomic:  gccaagttcagg-aa-agacg---g-gt-g--c-a------ga--gag-aa-a-aa----
          |||||||||||| || |||||   | ||  |  |      ||  ||| ||  | ||
cDNA:     gccaagttcaggTaaTagacgTTTgTgtTgTTcTaTTTTTTgaTTgagTaaTaTaaTTTT
Protein:   P  S  S  G  N  R  R  L  C  C  S  I  F  W  L  S  N  I  I  F

Genomic:  --a-g-a-ccg-------c-g--g-a---a---g-ag------ctcagcaa--a---a-
            | | | |||       | |  | |   |   | ||      ||||||||  |   |
cDNA:     TTaTgTaTccgTTTTTTcTgTgTaTTTaTTTTgTagTTTTTTctcagcaaTaTaTTaT
Protein:   Y  V  S  V  F  F  C  C  I  Y  F  V  V  F  S  Q  Q  L  F  I

Genomic:  ag--gaaTTTTTgg-gg-gg----a--gatt-gccggg-a--aaa-a--at-cgcg
          ||  |||     || || ||    |  |||| ||||||| |  ||| |  || ||||
cDNA:     agTTgaa----ggTggTggTTTTaTTgattTgccgggTaTTaaaTaTTatTcgcgTTTT
Protein:   V  E        G  G  G  F  I  D  L  P  G  I  K  Y  Y  S  R  F

Genomic:  ---a--agtgtttgtcagataatattaaacattgcttatattgct
          |||   |||||||||||||||||||||||||||||||||||||||
cDNA:     TTTaTaTagtgtttgtcagataatattaaacattgcttatattgct
Protein:   F  Y  *  C  L  S  D  N  I  K  H  C  L  Y  C
```

Legend:
- : Substitution
- : Insertion
- : Deletion
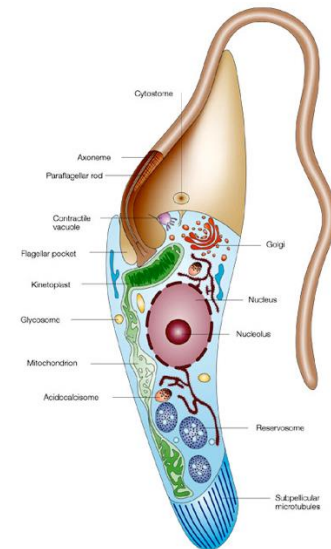
*Trypanosoma cruzi*

Gene *rps12* from the mitochondrial genome of *Crithidia fasciculata*
Genomic length: 209 bp
cDNA length: 338 bp
Editing type: U Insertion (events: 59)
Editing type: U Deletion (events: 3)
Total Number of editing events: 62

From the REDIdb database

# RNA editing by base substitution in plant mt

```
ACGCTAGAAGGTGCAAAATTAACTGGAGCGGGAGCCGCTACCATTGCTTCGGCGGGAGCTGCT
   T                        T                              T
 T>M L   E   G   A   K   L  T>I G   A   G   A   A   T   I   A  S>L A   G   A   A

GTAGGTATTGGAAACGCTCCCAGTCCTTCGATTCATGGAGTTGCGTGAAATCCACCATTAGCT
           T  TT     T     T                          C        T
 V   G   I   G   N  A>VP>F  S  P>SS>L I   H   G   V   A  *>R N   P  P>S  L   A

AAGCAATCATCTGGTTACGTTATTCCAGGTCCTGTTTCAACCGAAGCTACCGTTTCGCTTGCC
       T   T           C         T      TT  C   T                T   C   T  T
 K   Q  S>LS>F G   Y  V>A I  P>L  G  P>FV>AS>L T   E   A  T>IV>AS>LL>F A

CCAACGATGGCGTCTCCAATCCCATCCGTATCCCAA
   T   T           T   T       T   T       TTT
 P>LT>M M   A  S>FP>L I  P>LS>F V  S>FQ>*
```
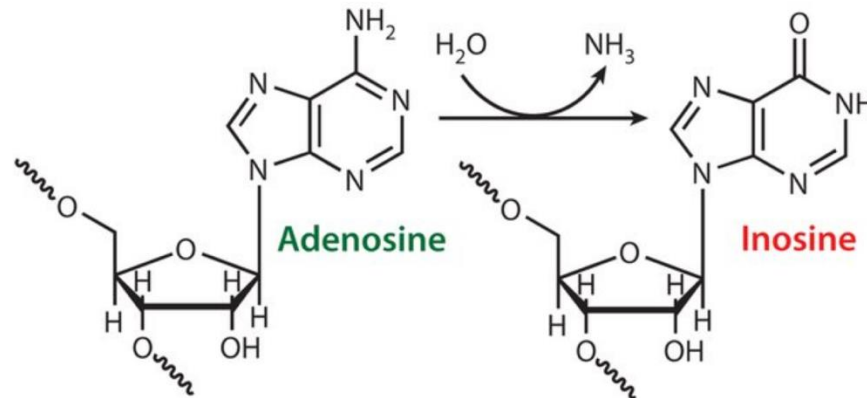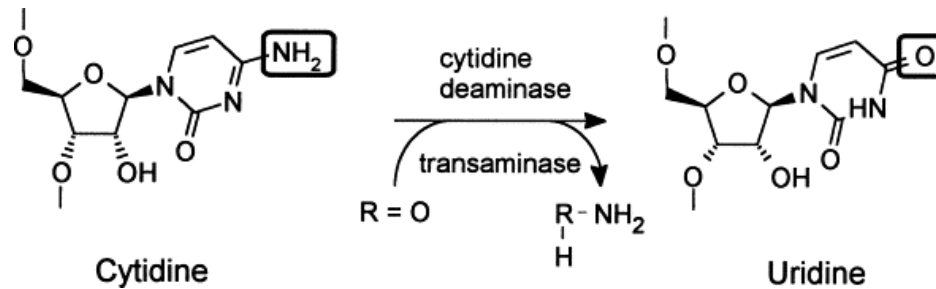
Gene *atp9* from the mitochondrial genome of *Isoetes engelmannii.*

Yellow and magenta shading indicate changes introduced through C-to-U or U-to-C editing events in the *atp9* mRNA.
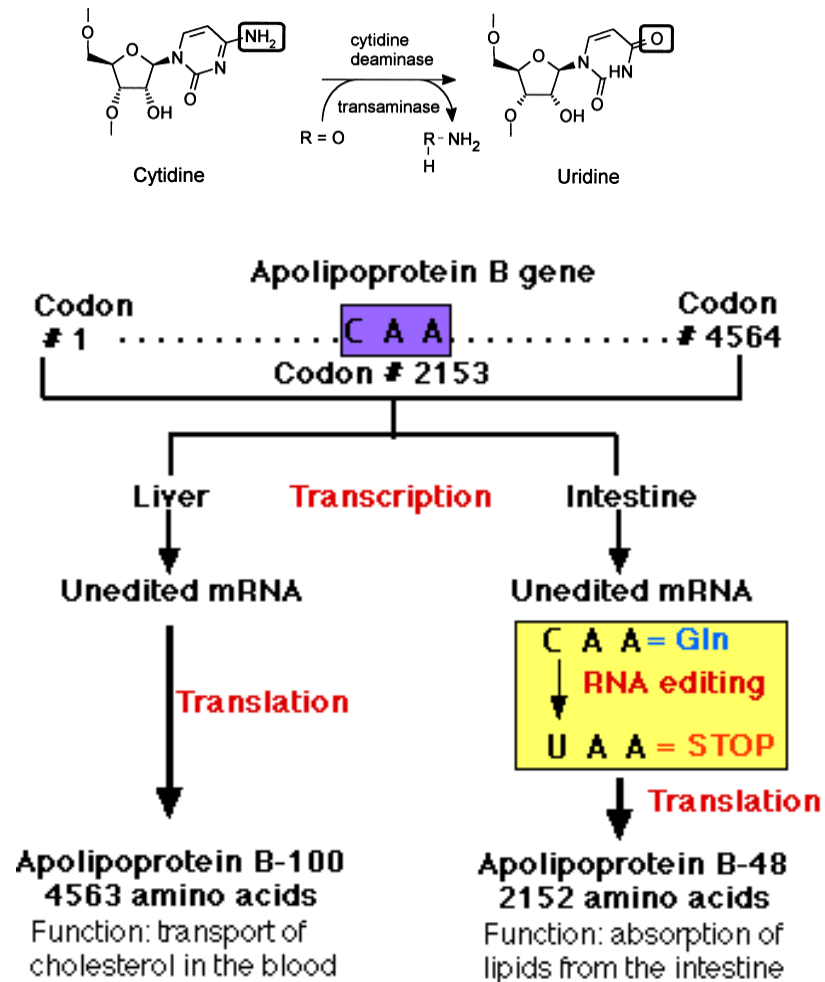
Grewe et al., 2009 NAR

# RNA editing in mammals

In mammals, RNA editing occurs in the nucleus and cytoplasm and mainly by C-to-U or A-to-I conversions.

# C-to-U RNA editing in humans

C-to-U editing is carried out by the APOBEC family of enzymes an only a few instances are known.



Cytidine — cytidine deaminase / transaminase (R = O, R-NH₂) — Uridine
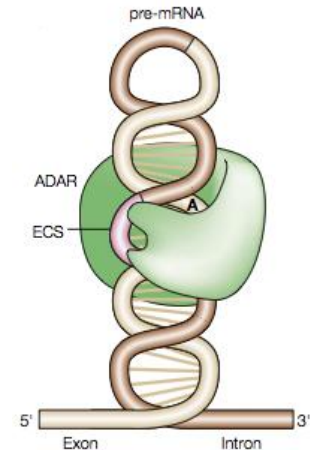


Apolipoprotein B gene
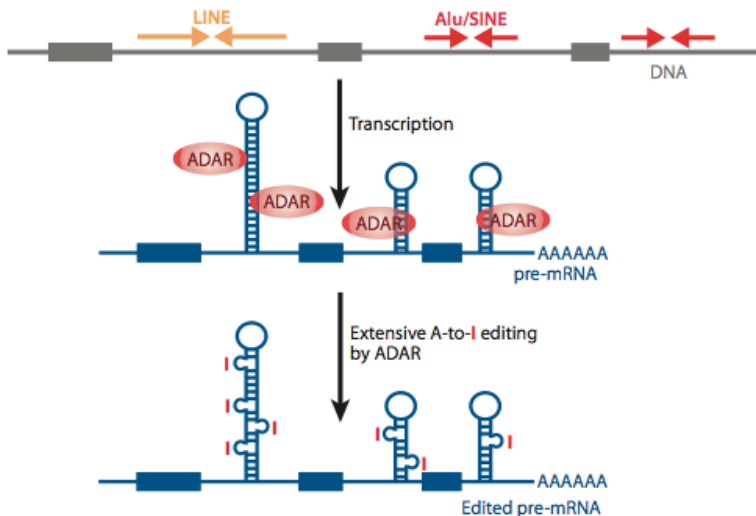
# A-to-I RNA editing in humans

A-to-I is carried out by ADAR (adenine deaminases acting on RNA) enzymes. ADAR1 and ADAR2 are expressed in almost all human tissues. ADAR3, instead, is expressed in the brain only. ADAR enzymes can deaminate As included in RNA duplexes.
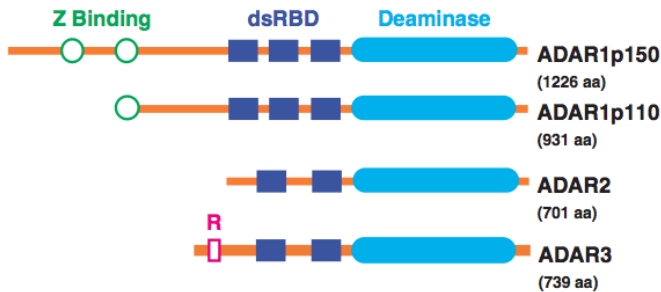


Zinshteyn et al., 2009, WIREs Syst Biol Med



Keegan et al., 2001, Nature Rev. Gen.



Nishikura 2010 Annu. Rev. Biochem

Almost all RNA editing sites identified in the human transcriptome reside in non-coding regions that consist of inversely oriented repetitive elements, mostly Alu repeats (~90%) and some LINE repeats (~10%), representing ~13% and ~21% of the human genome, respectively.

# Effects of A-to-I RNA editing

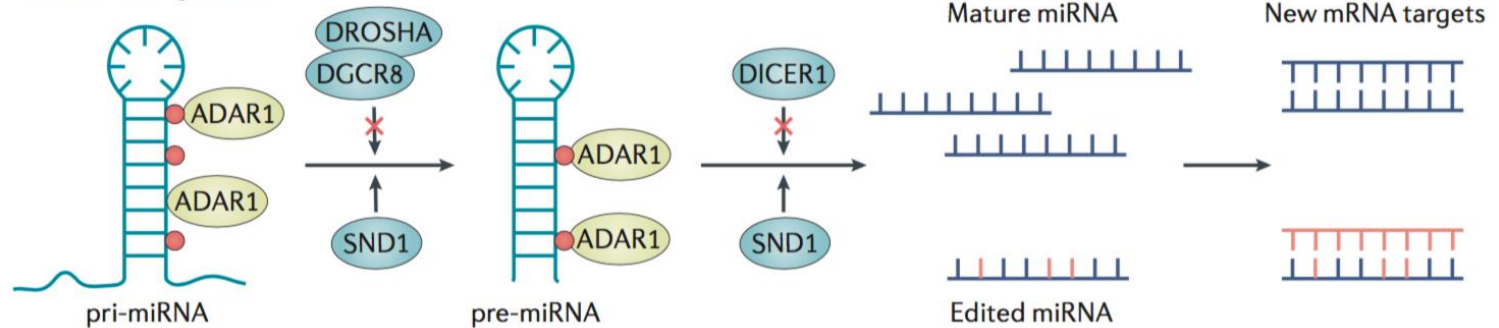A-to-I editing can modulate the gene expression at different levels.
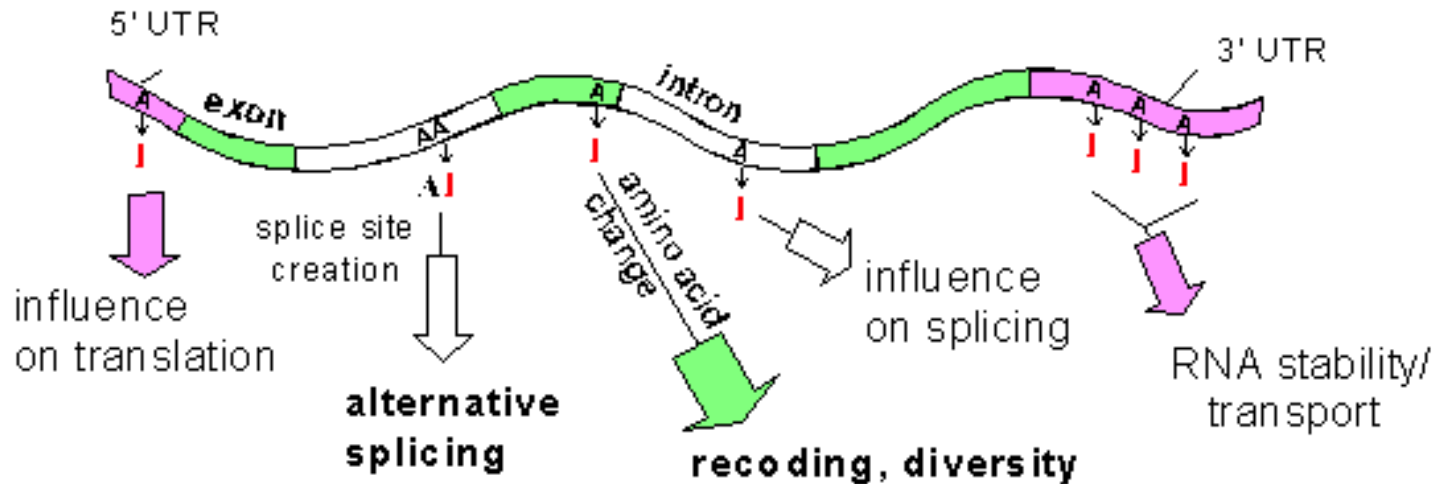


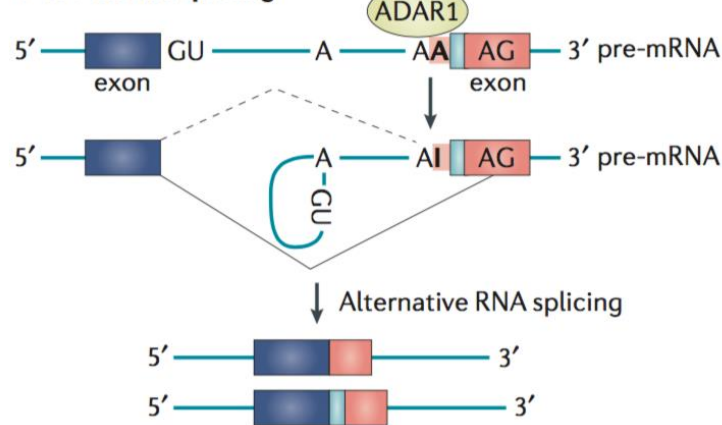NuclearA-to-I editing: known (**bold**) and proposed functions.

Maas S. web site
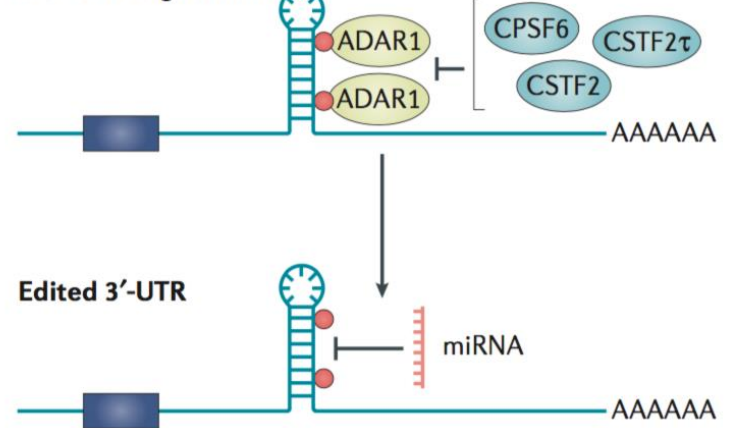
# Effects of RNA editing in mammals



Consequences of RNA editing by ADAR. a | Adenosine-to-inosine (A-to-I) editing-dependent regulation of microRNA (miRNA) biogenesis and targeting. b | RNA editing at alternative 3'-acceptor sites converts the intronic AA into AI dinucleotides, which mimic the conserved AG sequences normally found at 3'-splicing sites. This introduction of a new splicing acceptor site results in an alternatively spliced mRNA with an insertion (light blue). c | ADAR1 competes with canonical 3'-untranslated region (UTR) processing. Moreover, RNA editing changes within the miRNA target sequences may prevent miRNA binding and therefore downregulation of the mRNA.

Jiang et al. 2017 Nat Rev Cancer

# A-to-I RNA editing in the innate immune response



Walkley and Li 2017 Genome Biol
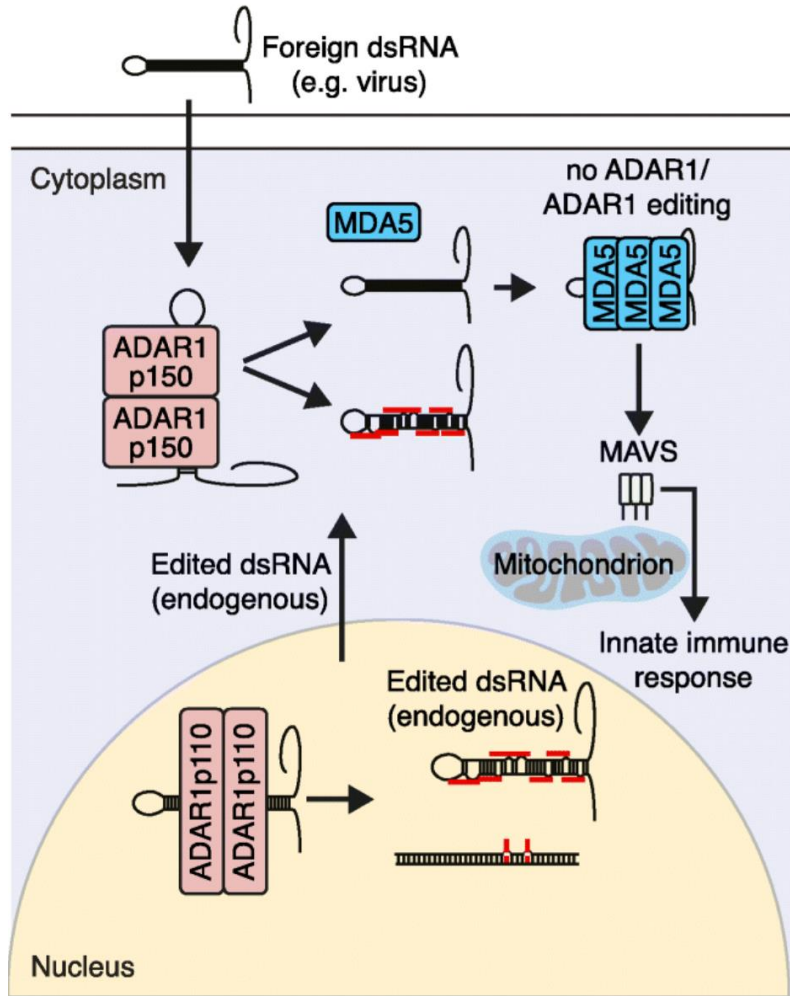
ADAR1 is present in the nucleus (ADAR1 p110) and cytoplasm (ADAR1 p150) and can edit endogenous RNA. ADAR1 is required to edit endogenous RNA to prevent the activation of the cytosolic pattern recognition receptor MDA5 in the cytosol, leading to induction of the innate immune/interferon response. ADAR1 can also edit viral dsRNA and participate in the innate immune response as a direct interferon-stimulated gene (ADAR1 p150 isoform). The absence of ADAR1 or the absence of ADAR1-mediated editing leads to innapropriate activation of the MDA5–MAVS axis.

# Recoding RNA editing in humans

RNA editing deregulation is associated to several human diseases including neurological disorders as major depression, schizophrenia, epilepsy, amyotrophic lateral sclerosis (ALS) and cancer.

# Recoding RNA editing in humans



**Differential levels of RNA editing in Hippocampus**
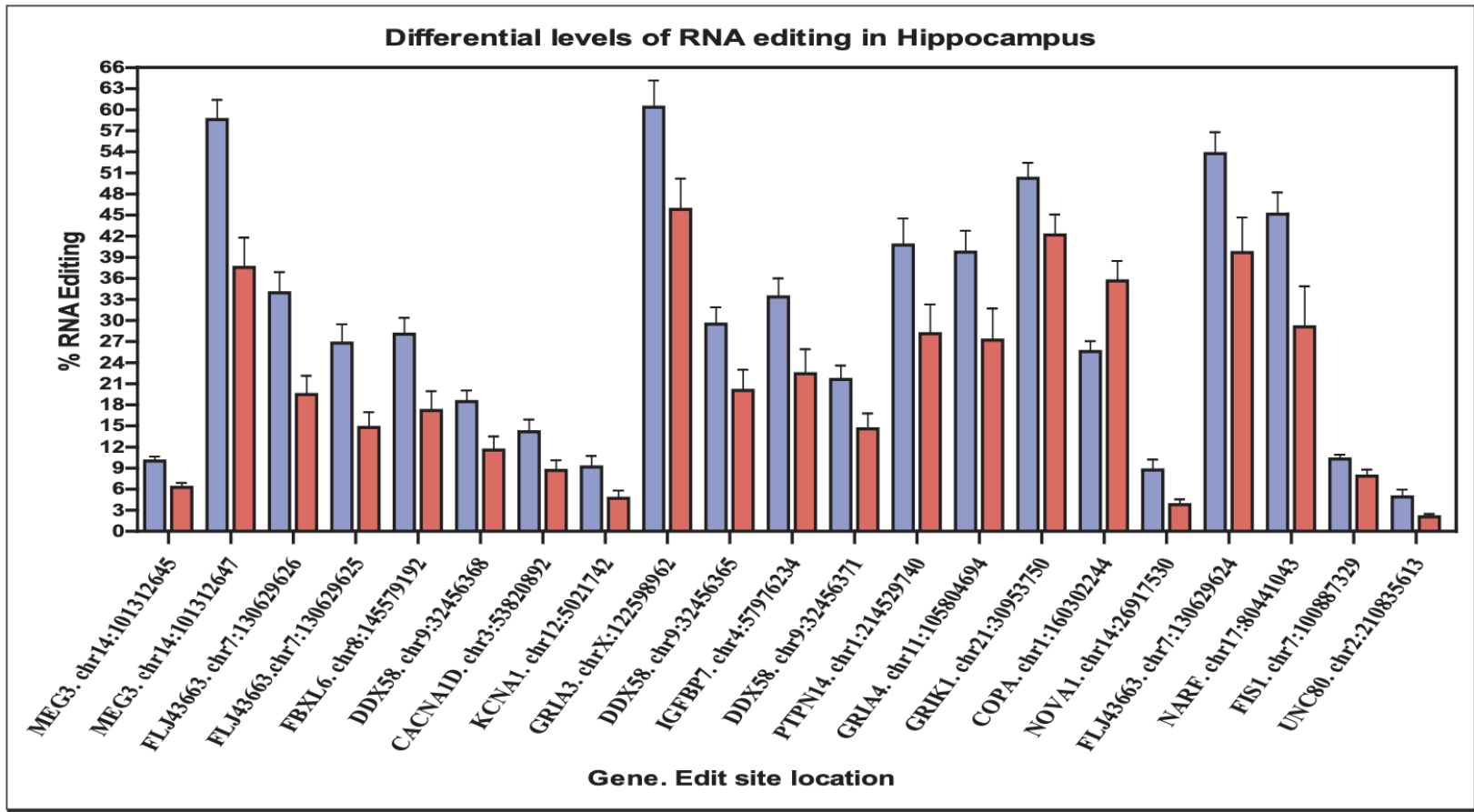
RNA editing target sites that show significant differential levels in NDC (non demended controls) and AD (Alzheimer samples). In Hippocampus, 21 editing target sites exhibiting differential levels of A-to-I RNA editing were detected in AD (red bars) and in NDC (blue bars). Twenty sites show hypo-editing in AD and one site (COPA. Chr1:160302244) shows hyper-editing (n = 28 for AD, n = 20 in NDC)

KHERMESH et al. 2016 RNA

# RNA editing during neuronal differentiation

RNA Editing is differentially regulated during neuronal differentiation and maturation. Neuronal induction driving self-renewing stem cells into neural progenitor cells involves editing of the Q/R site in the mRNA encoding the excitatory AMPA-type glutamate receptor subunit, GluA2. This re-coding editing event is rapidly increased in the early neuronal lineage. Selected editing substrates for important neurotransmitters are: the R/G site of GluA2, the I/M site of Gabra-3, and the A and B sites in the serotonin (5HT2C) receptor. Editing levels increase with different efficiencies from embryogenesis to adulthood.



Behm and Ohman 2016
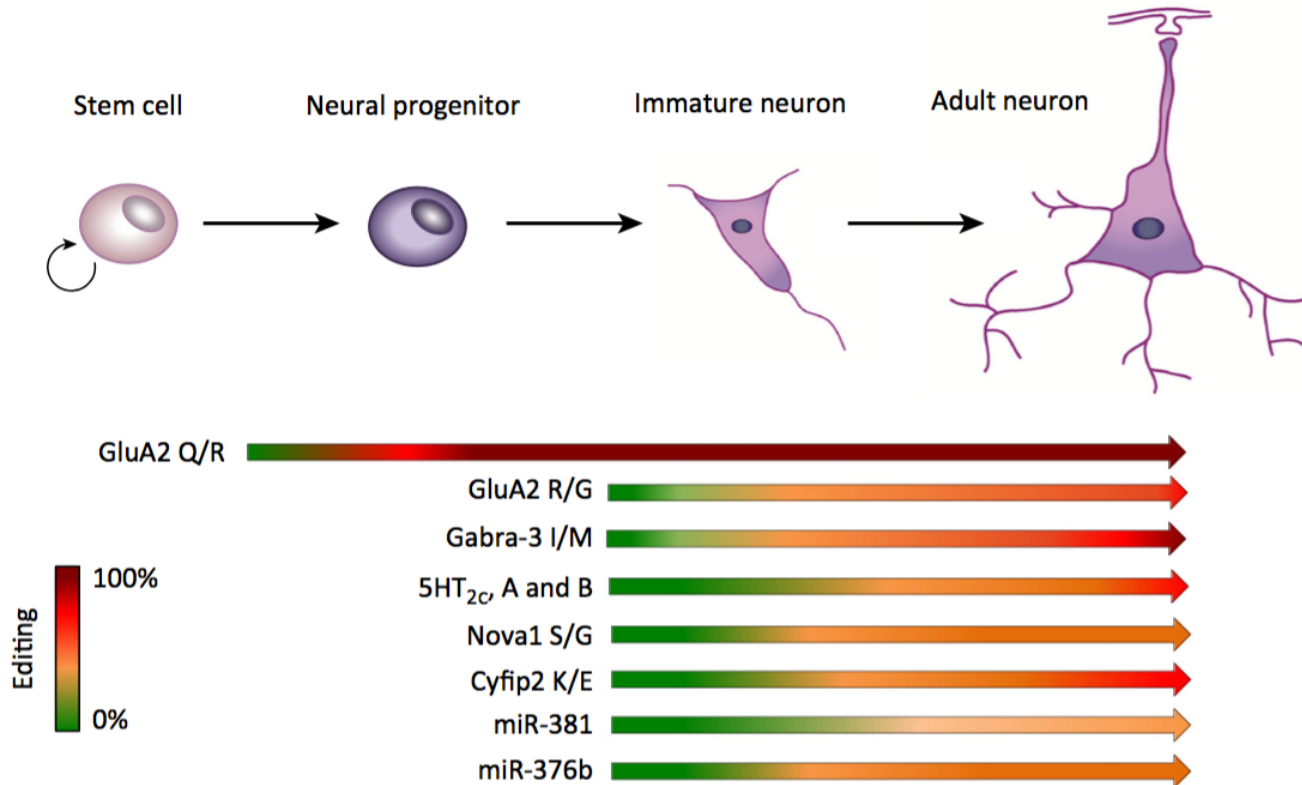
# RNA editing: experimental detection

RNA editing changes can be experimentally detected by comparing the genomic *locus* with the corresponding cDNA sequenced by classical Sanger methodology.
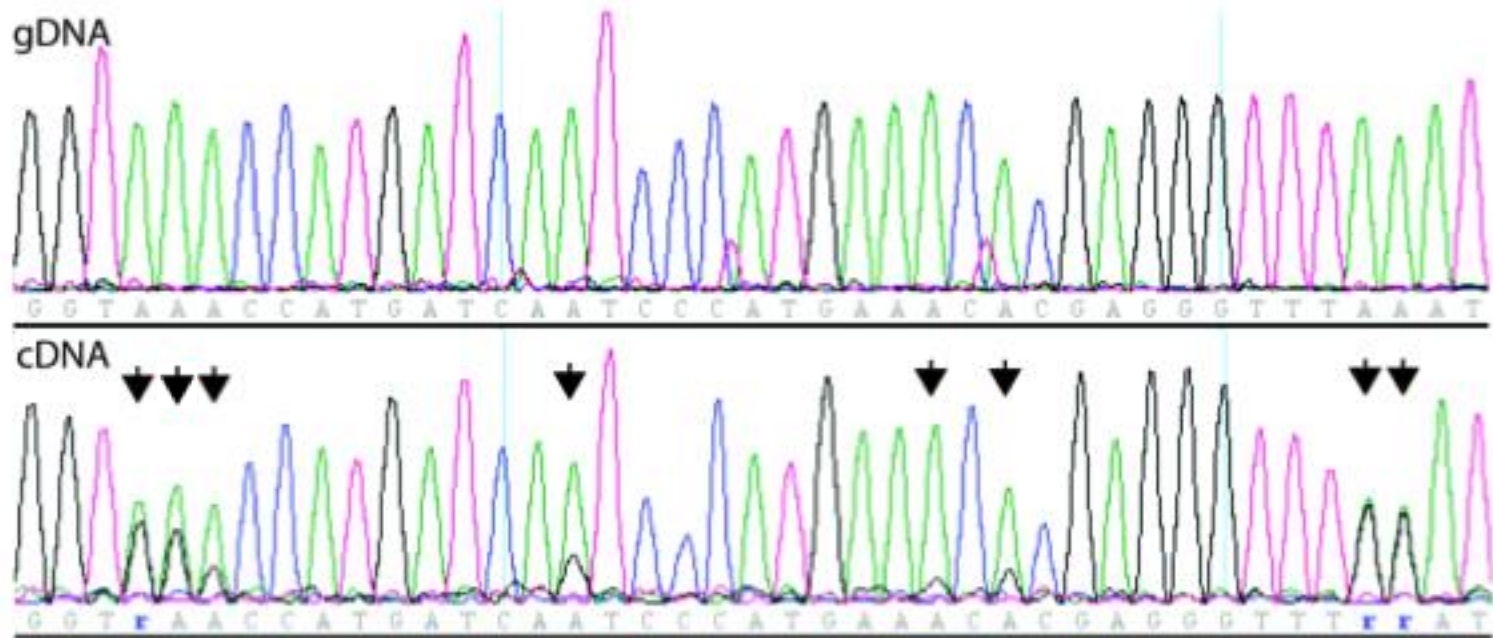


Li et al. 2009

Editing sites (indicated by black arrows) in the amyloid beta A4 precursor protein-binding (*APBA1*) gene in cerebellum.

# RNA editing: bioinformatics

The detection of RNA editing in human by conventional techniques is not feasible for large-scale experiments. Therefore, many candidate events have been mainly identified by computational analyses, employing mRNA/EST alignments onto the genome of origin.
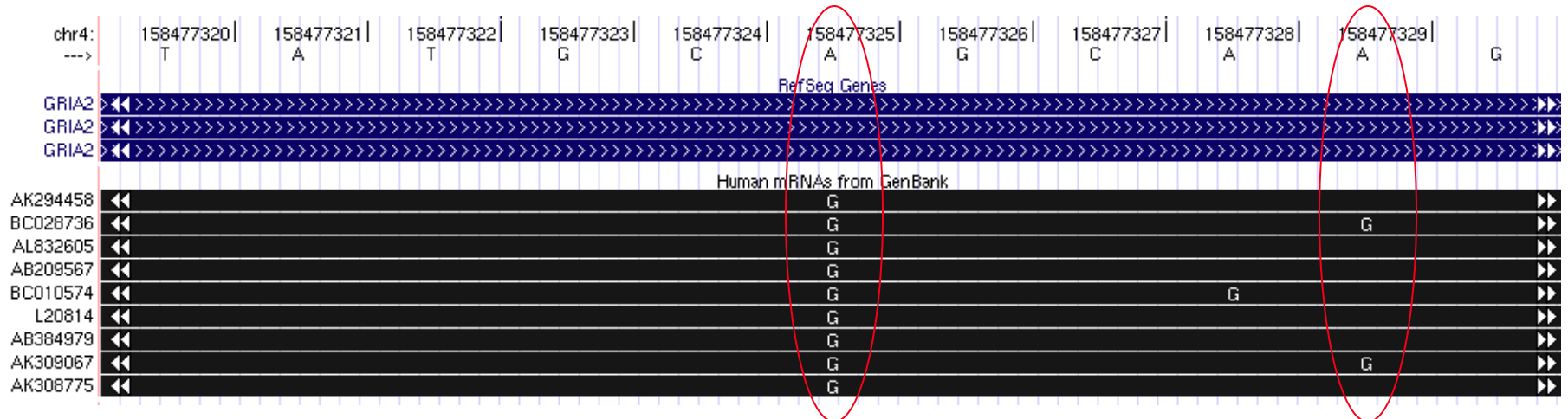
# The impact of RNA editing on scientific community

RNA editing publications since 1986
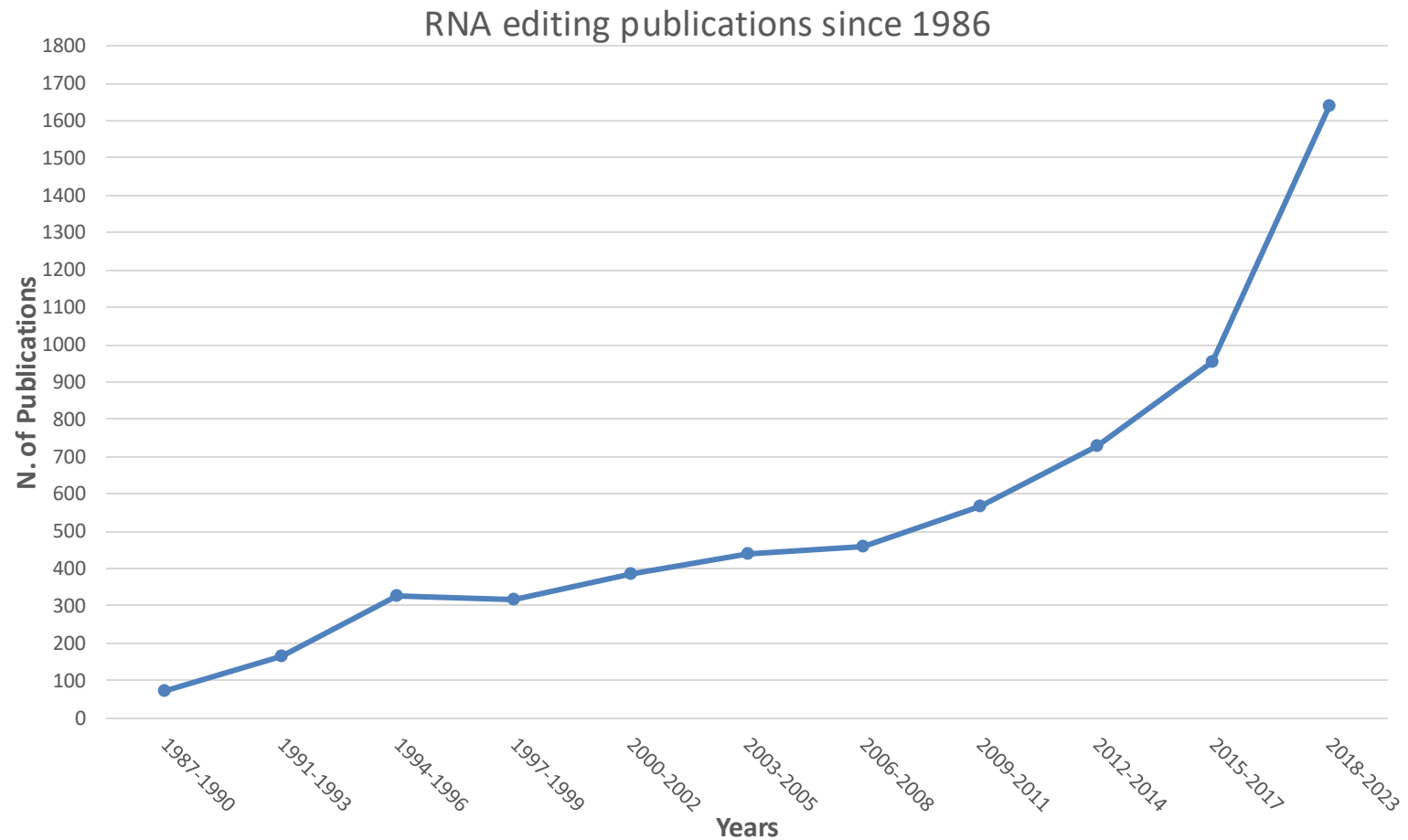


Data from PubMed (last update 10/01/2024)

# RNA editing and NGS

Massive RNA sequencing can facilitate the study of entire transcriptomes as well as post-transcriptional events occurring herein as alternative splicing and RNA editing.

**Genome**

**Short reads**

**Genome**

**ESTs/cDNAs**

**Few clones are sequenced by Sanger method and differences between the genome and the related cDNAs are scored as RNA editing sites**
**Each editing site is supported by a very restricted number of transcripts.**
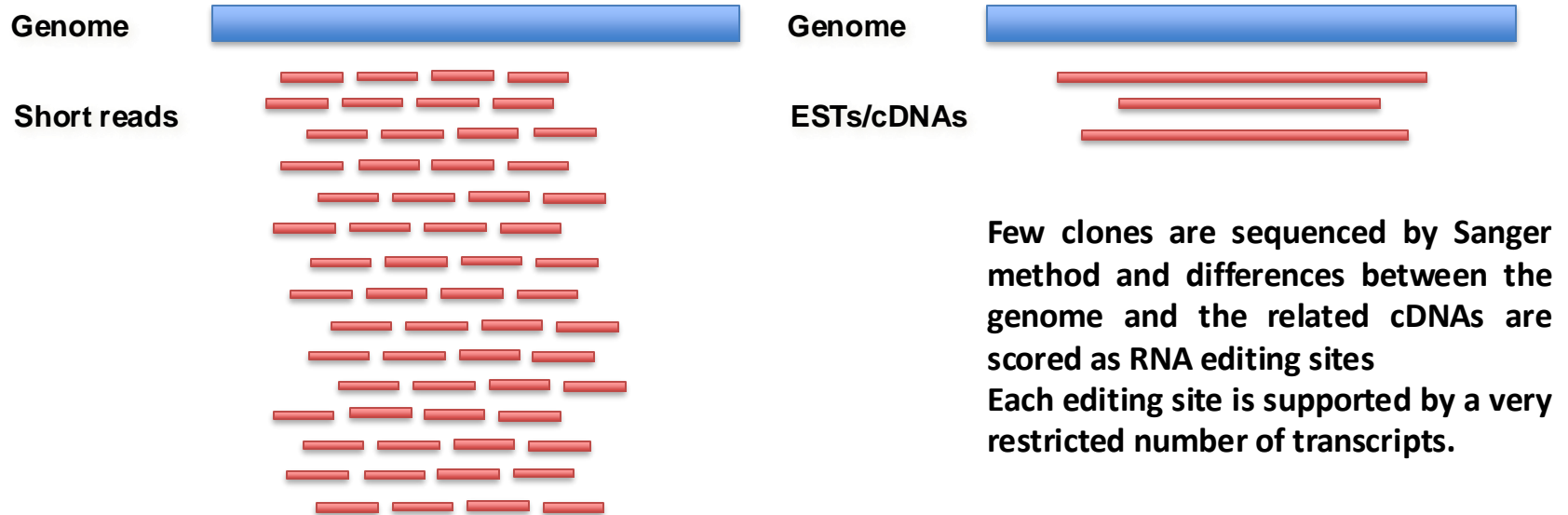
**Using NGS, each genomic position can be supported by a large number of sequences and this can greatly improve the detection of RNA editing substitutions.**

**Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing**

Ernesto Picardi[1], David S. Horner[2], Matteo Chiara[2], Riccardo Schiavon[3], Giorgio Valle[3] and Graziano Pesole[1,4,*]

# RNA editing detection by NGS data

Since Vitis, RNA-Seq data have been employed to identify RNA editing changes in human transcriptomes. However, recent methods have been mainly applied to cell lines like lymphoblastoid lines which are not the elective material for RNA editing investigations in human.

**Sciencexpress** — **Research Article**

## Widespread RNA and DNA Sequence Differences in the Human Transcriptome

Mingyao Li,[1]* Isabel X. Wang,[8]* Yun Li,[6,7] Alan Bruzel,[8] Allison L. Richards,[4] Jonathan M. Toung,[5] Vivian G. Cheung[2,3,8†]

## Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome

Zhiyu Peng[1,5], Yanbing Cheng[1,5], Bertrand Chin-Ming Tan[2,5], Lin Kang[1], Zhijian Tian[1], Yuankun Zhu[1], Wenwei Zhang[1], Yu Liang[1], Xueda Hu[1], Xuemei Tan[1], Jing Guo[1], Zirui Dong[1], Yan Liang[1], Li Bao[1] & Jun Wang[1,3,4]

**RNA** — A PUBLICATION OF THE RNA SOCIETY

### RNA editing of protein sequences: A rare event in human transcriptomes

Claudia L. Kleinman, Véronique Adoue and Jacek Majewski

RNA published online July 25, 2012
Access the most recent version at doi:10.1261/rna.033233.112

## Accurate Identification of A-to-I RNA editing in human by transcriptome sequencing

Jae Hoon Bahn, Jae Hyung Lee, Gang Li, et al.

Genome Res. published online September 29, 2011
Access the most recent version at doi:10.1101/gr.124107.111

**nature genetics**

## Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals

Young Seok Ju[1,2,9], Jong-Il Kim[1,3–5,9], Sheehyun Kim[1,2,9], Dongwan Hong[1,8], Hansoo Park[1,6], Jong-Yeon Shin[1,5], Seungbok Lee[1,4], Won-Chul Lee[1,4], Sujung Kim[5], Saet-Byeol Yu[5], Sung-Soo Park[5], Seung-Hyun Seo[5], Ji-Young Yun[5], Hyun-Jin Kim[1,4], Dong-Sung Lee[1,4], Maryam Yavartanoo[1,4], Hyunseok Peter Kang[1], Omer Gokcumen[6], Diddahally R Govindaraju[6], Jung Hee Jung[2], Hyonyong Chong[2,7], Kap-Seok Yang[2], Hyungtae Kim[2], Charles Lee[6] & Jeong-Sun Seo[1–5,7]

## Accurate identification of human *Alu* and non-*Alu* RNA editing sites

Gokul Ramaswami[1,3], Wei Lin[2,3], Robert Piskol[1,3], Meng How Tan[1], Carrie Davis[2] & Jin Billy Li[1]

**CSH PRESS** — **GENOME RESEARCH**

**A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes**

# RNA-Seq: analysis workflow
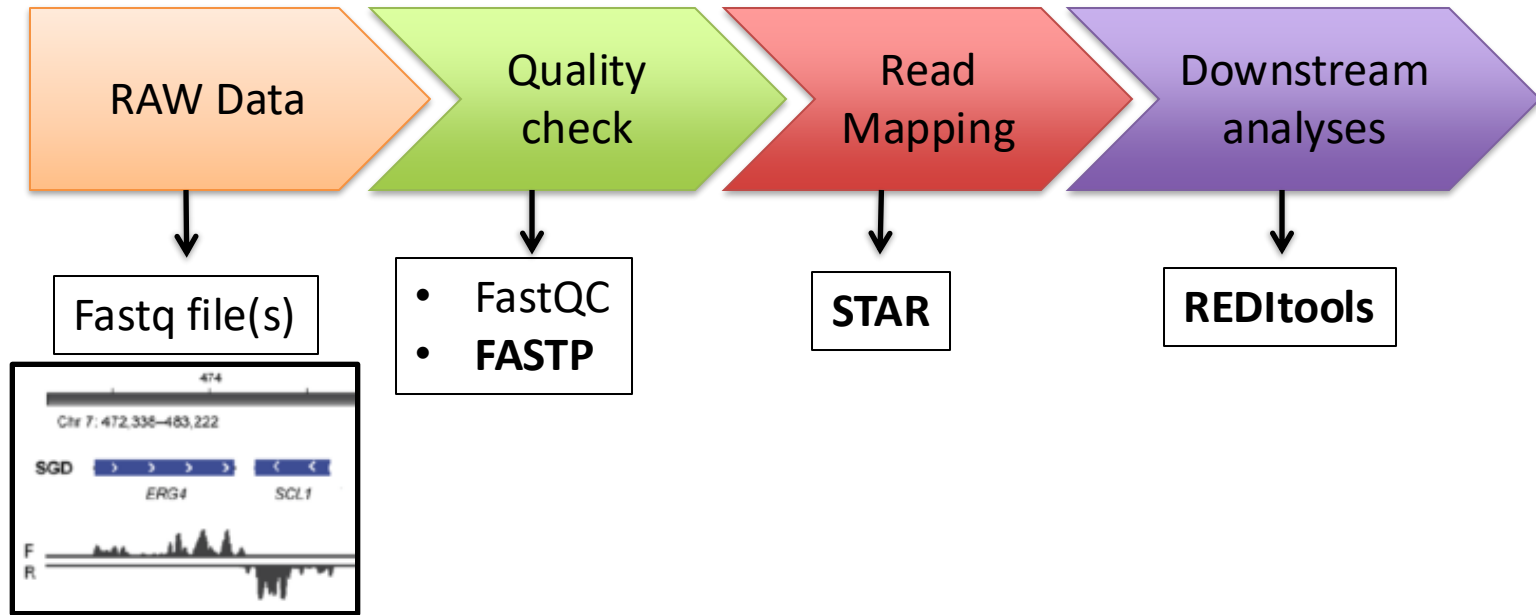


Bioinformatics workflow to preprocess data.

Reliable RNA editing calls require good quality WGS and RNAseq reads. Raw reads in fastq format are quality checked using FASTQC and cleaned using FASTP. Then, reads are aligned to the reference genome using a splice-aware software like STAR. Aligned reads are converted into the standard BAM format for the downstream detection of RNA editing.

(WGS reads are mapped using BWA)

Lo Giudice et al. 2020 Nat. Prot.
Lo Giudice et al. 2020 Front. In Gen.

# RNA-Seq: analysis workflow

To demonstrate the effects of pre-processing and genome alignment steps on RNA editing calling, we used a single GTEx RNAseq experiment from human cerebellum (run accession SRR607967). Two datasets were generated, the first containing original raw reads and the second including trimmed reads (by fastp). Both datasets were aligned onto the hg19 and hg38 reference genomes using three different aligners, BWA designed for unspliced reads and STAR and HISAT2 optimized for handling spliced reads. Resulting multi-alignments were processed with REDItools in order to provide the distribution of single RNA variants according to a common basic filtering scheme.



Lo Giudice et al. 2020
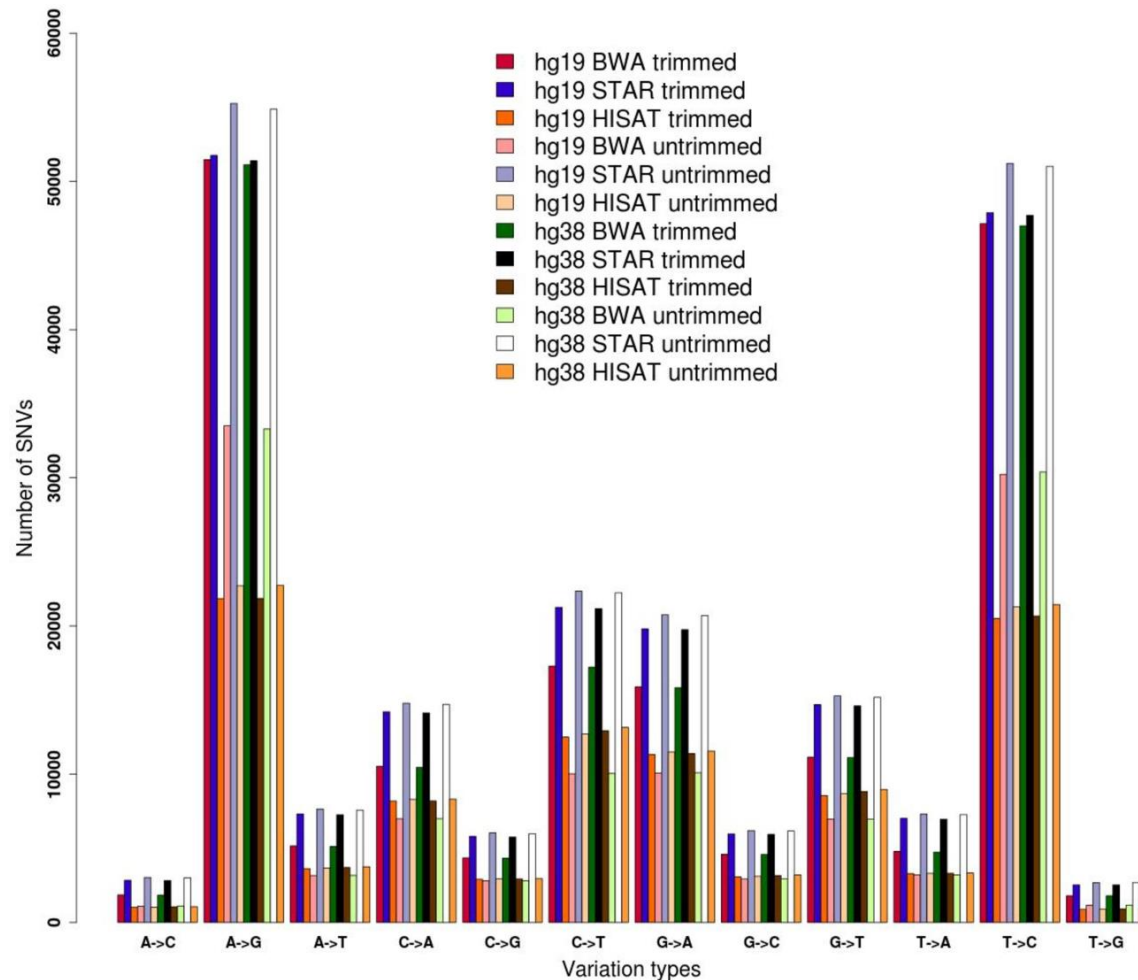
# RNA-Seq: RNA editing detection

We can employ NGS data (RNA-Seq, genome resequencing and exome sequencing) to study RNA editing at different levels:

✓ genome/exome Vs RNA-Seq to identify new events (REDItools);

*BIOINFORMATICS* **APPLICATIONS NOTE** *2013, pages 1–2*
*doi:10.1093/bioinformatics/btt287*

*Sequence analysis* — Advance Access publication June 5, 2013

**REDItools: high-throughput RNA editing detection made easy**
Ernesto Picardi[1,2,*] and Graziano Pesole[1,2,3,*]

✓ RNA-Seq to explore the presence of known A-to-I conversions;

*BIOINFORMATICS* **APPLICATIONS NOTE** *2011, pages 1–2*
*doi:10.1093/bioinformatics/btr117*

*Genome analysis*

**ExpEdit: a webserver to explore human RNA editing in RNA-Seq experiments**
Ernesto Picardi[1], D'Antonio Mattia[2], Danilo Carrabino[2], Tiziana Castrignanò[2] and Graziano Pesole[1,3,*]

✓ RNA-Seq to detect *de novo* new editing candidates;

OPEN ACCESS Freely available online — PLOS | ONE

**A Novel Computational Strategy to Identify A-to-I RNA Editing Sites by RNA-Seq Data: De Novo Detection in Human Spinal Cord Tissue**

Ernesto Picardi[1,2], Angela Gallo[3], Federica Galeano[3], Sara Tomaselli[3], Graziano Pesole[1,2,*]

1 Dipartimento di Bioscienze, Biotecnologie e Scienze Farmacologiche, Università di Bari, Bari, Italy, 2 Istituto di Biomembrane e Bioenergetica, Consiglio Nazionale delle Ricerche, Bari, Italy, 3 RNA Editing Laboratory, Oncohaematology Department, Ospedale Pediatrico "Bambino Gesù", IRCCS, Rome, Italy



```
          r1  GGGTGCCTTTATGCAGCAAGGATGCGATATT
          r2  GGGTGTCTTTATGCAGCAAGGATGCGATACTTCGC
Exome     r3  GGGTGCCTTTATGCAGCAAGGATGCGATATTTCG
          r4  GGGTGCCTTTATGCAGCAAGGATGCGATATTTCG
          r5  GGGTGCCTTTATGCAGCAAGGATGCGATATTTCG
              ...............A......................
     gDNA  TGGGTGCCTTTATGCAGCAAGGATGCGATATTTCGCC
              ...............G......................
          r1  GGGTGCCTTTATGCGGCAAGGATGCGATATT
          r2  GGGTGTCTTTATGCAGCAAGGATGCGATACTTCGC
RNA-Seq   r3  GGGTGCCTTTATGCGGCAAGGATGCGATATTTCG
          r4  GGGTGCCTTTATGCGGCAAGGATGCGATATTTCG
          r5  GGGTGCCTTTATGCGGCAAGGATGCGATATTTCG
```

| | | | | | | Nucleotide | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Location | Position | Reference base | Strand | Gene | Region | As | Cs | Gs | Ts | Coverage | Editing extent | Source |
| chr1 | 6081149 | A | + | KCNAB2 | CDS | 0 | 0 | 31 | 0 | 31 | 1.000 | D |
| chr4 | 158477325 | A | + | GRIA2 | CDS | 0 | 0 | 11 | 0 | 11 | 1.000 | D |
| chr5 | 150619602 | A | + | GM2A | CDS | 0 | 0 | 16 | 0 | 16 | 1.000 | D |
| chr5 | 150619632 | A | + | GM2A | CDS | 0 | 0 | 12 | 0 | 12 | 1.000 | D |
| chr11 | 62214851 | A | - | BSCL2 | CDS | 0 | 0 | 20 | 0 | 20 | 1.000 | D |
| chr16 | 57102927 | A | + | NDRG4 | CDS | 0 | 0 | 93 | 0 | 93 | 1.000 | D |
| chr11 | 77468301 | A | - | NDUFC2 | CDS | 9 | 0 | 17 | 0 | 26 | 0.654 | D |

http://epigen.hpc.cineca.it/expedit/index.php

```
     gDNA  AGCTGGCCAGATACATTAAGACCAGTGCTCACTATGAAG
              ...............G......................
          r1  GCTGGCCAGATACATTGAGACCAGTGCTCAC
          r2  GCTGGCCAGATACATTAAGACCAGTGCTCAC
          r3  CTGGCCAGATACATTGAGACCAGTGCTCACTATGAAG
RNA-Seq   r4  CTGGCCAGATACATTGAGACCAGTGCTCACTATG
          r5  CTGGCCAGATACATTAAGACCAGTGCTCACTATGAAG
          r6  CTGGCCAGATACATTAAGACCAGTGCTCACTATGAAG
          r7  CTGGCCAGATACATTGGGACCAGTGCTCACTATGAAG
          r8  CTGGCCAGATACATTGAGACCAGTGCTCACT
          r9  CTGGCCAGATACATTGAGACCAGTGCTCACTATGAAG
```

# REDItools

"REDItools" are a suite of python scripts to investigate RNA editing at large-scale employing RNA-Seq as well as DNA-Seq (WGS/WES) massive data.
Starting point is a BAM file of aligned reads onto the reference genome.



**REDItoolDnaRna.py**

RNA-Seq and
DNA-Seq

**REDItoolKnown.py**

RNA-Seq and
known events

**REDItoolDenovo.py**

RNA-Seq only

**https://github.com/BioinfoUNIBA**
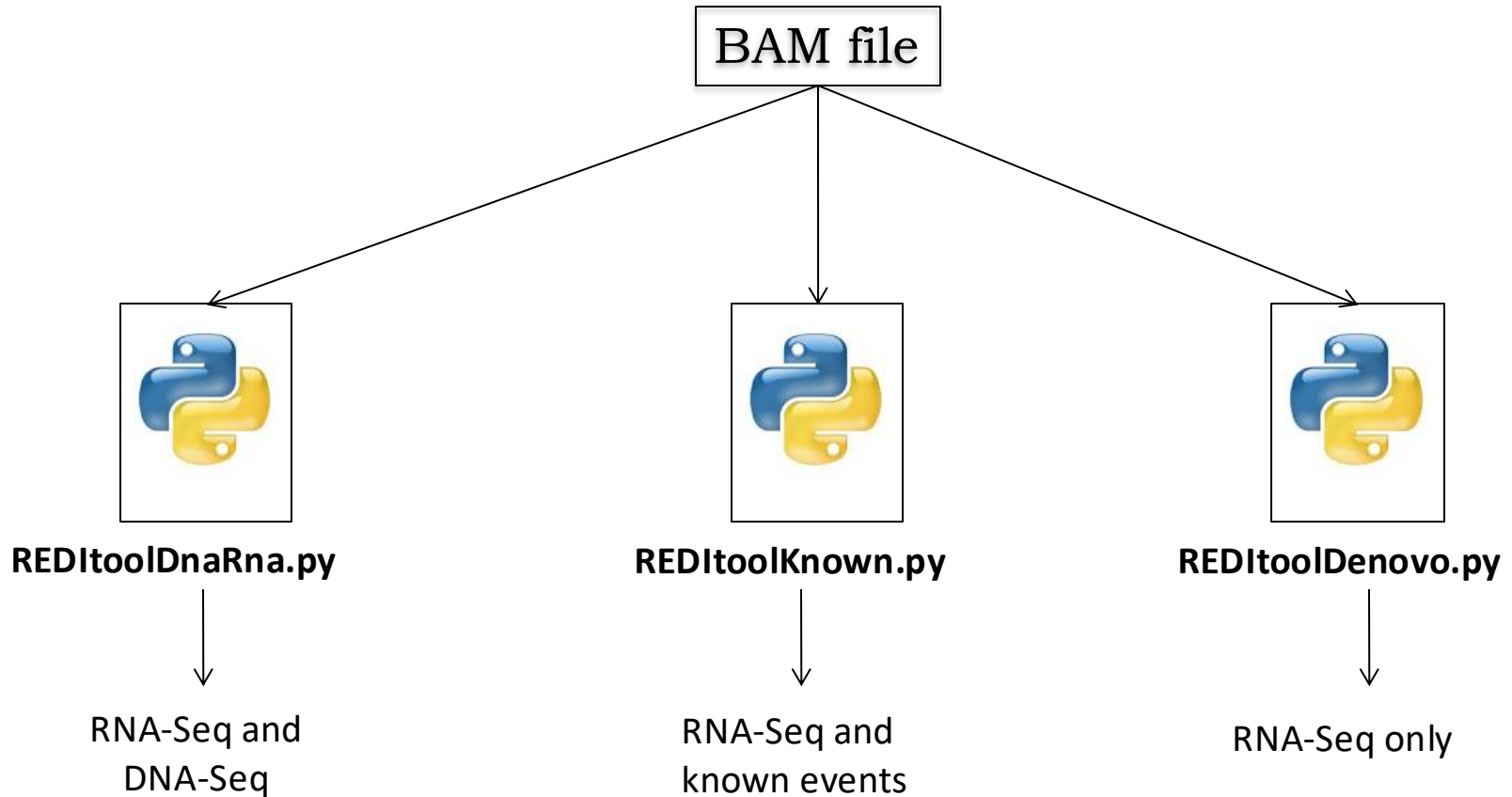
Picardi and Pesole 2013 Bioinformatics

# RNA editing and NGS

Workflow to call RNA editing by REDItools.

BAM file

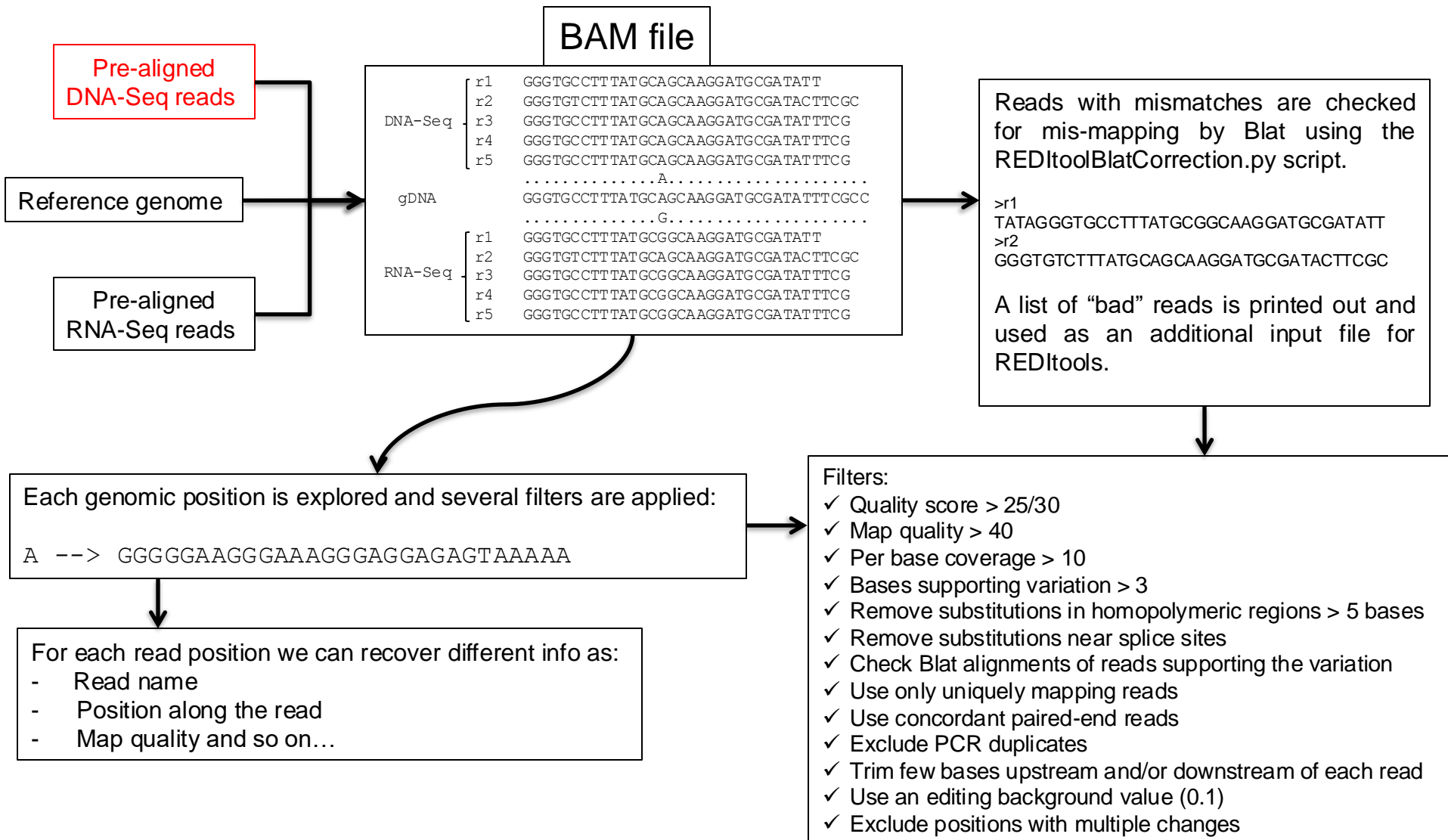| | | |
|---|---|---|
| DNA-Seq | r1 | GGGTGCCTTTATGCAGCAAGGATGCGATATT |
| | r2 | GGGTGTCTTTATGCAGCAAGGATGCGATACTTCGC |
| | r3 | GGGTGCCTTTATGCAGCAAGGATGCGATATTTCG |
| | r4 | GGGTGCCTTTATGCAGCAAGGATGCGATATTTCG |
| | r5 | GGGTGCCTTTATGCAGCAAGGATGCGATATTTCG |

```
..............A.....................
gDNA    GGGTGCCTTTATGCAGCAAGGATGCGATATTTCGCC
..............G.....................
```

| | | |
|---|---|---|
| RNA-Seq | r1 | GGGTGCCTTTATGCGGCAAGGATGCGATATT |
| | r2 | GGGTGTCTTTATGCAGCAAGGATGCGATACTTCGC |
| | r3 | GGGTGCCTTTATGCGGCAAGGATGCGATATTTCG |
| | r4 | GGGTGCCTTTATGCGGCAAGGATGCGATATTTCG |
| | r5 | GGGTGCCTTTATGCGGCAAGGATGCGATATTTCG |

Pre-aligned DNA-Seq reads

Reference genome

Pre-aligned RNA-Seq reads

Reads with mismatches are checked for mis-mapping by Blat using the REDItoolBlatCorrection.py script.

```
>r1
TATAGGGTGCCTTTATGCGGCAAGGATGCGATATT
>r2
GGGTGTCTTTATGCAGCAAGGATGCGATACTTCGC
```

A list of "bad" reads is printed out and used as an additional input file for REDItools.

Each genomic position is explored and several filters are applied:

```
A --> GGGGGAAGGGAAAGGGAGGAGAGTAAAAA
```

For each read position we can recover different info as:
- Read name
- Position along the read
- Map quality and so on…

Filters:
- ✓ Quality score > 25/30
- ✓ Map quality > 40
- ✓ Per base coverage > 10
- ✓ Bases supporting variation > 3
- ✓ Remove substitutions in homopolymeric regions > 5 bases
- ✓ Remove substitutions near splice sites
- ✓ Check Blat alignments of reads supporting the variation
- ✓ Use only uniquely mapping reads
- ✓ Use concordant paired-end reads
- ✓ Exclude PCR duplicates
- ✓ Trim few bases upstream and/or downstream of each read
- ✓ Use an editing background value (0.1)
- ✓ Exclude positions with multiple changes

# RNA editing and NGS

| chr1 | 9327108 | A | 2 | 50 | 38.22 | [47, 0, 3, 0] | AG | 0.06 | 34 | 36.21 | [34, 0, 0, 0] | - | 0.00 |
| chr1 | 9327117 | A | 2 | 48 | 38.67 | [46, 0, 2, 0] | AG | 0.04 | 36 | 36.56 | [36, 0, 0, 0] | - | 0.00 |
| chr1 | 9327140 | A | 2 | 66 | 38.24 | [65, 0, 1, 0] | AG | 0.02 | 31 | 37.29 | [31, 0, 0, 0] | - | 0.00 |
| chr1 | 9327141 | A | 2 | 64 | 38.20 | [63, 0, 1, 0] | AG | 0.02 | 30 | 36.30 | [30, 0, 0, 0] | - | 0.00 |
| chr1 | 9328027 | A | 2 | 33 | 37.79 | [29, 0, 4, 0] | AG | 0.12 | 18 | 37.89 | [18, 0, 0, 0] | - | 0.00 |
| chr1 | 9328047 | A | 2 | 42 | 38.55 | [35, 0, 7, 0] | AG | 0.17 | 17 | 39.53 | [17, 0, 0, 0] | - | 0.00 |
| chr1 | 9328052 | A | 2 | 43 | 38.74 | [42, 0, 1, 0] | AG | 0.02 | 17 | 38.71 | [17, 0, 0, 0] | - | 0.00 |

REDItools traverse RNAseq (and WGS) multiple read alignments position by position looking at RNA mismatches (supported by WGS reads without SNP evidence). Several filters are applied to each position, and resulting variants are outputted in tab-delimited tables.

# RNA editing and NGS
## Effect of filters

Filters can directly affect the number and quality of RNA variants.
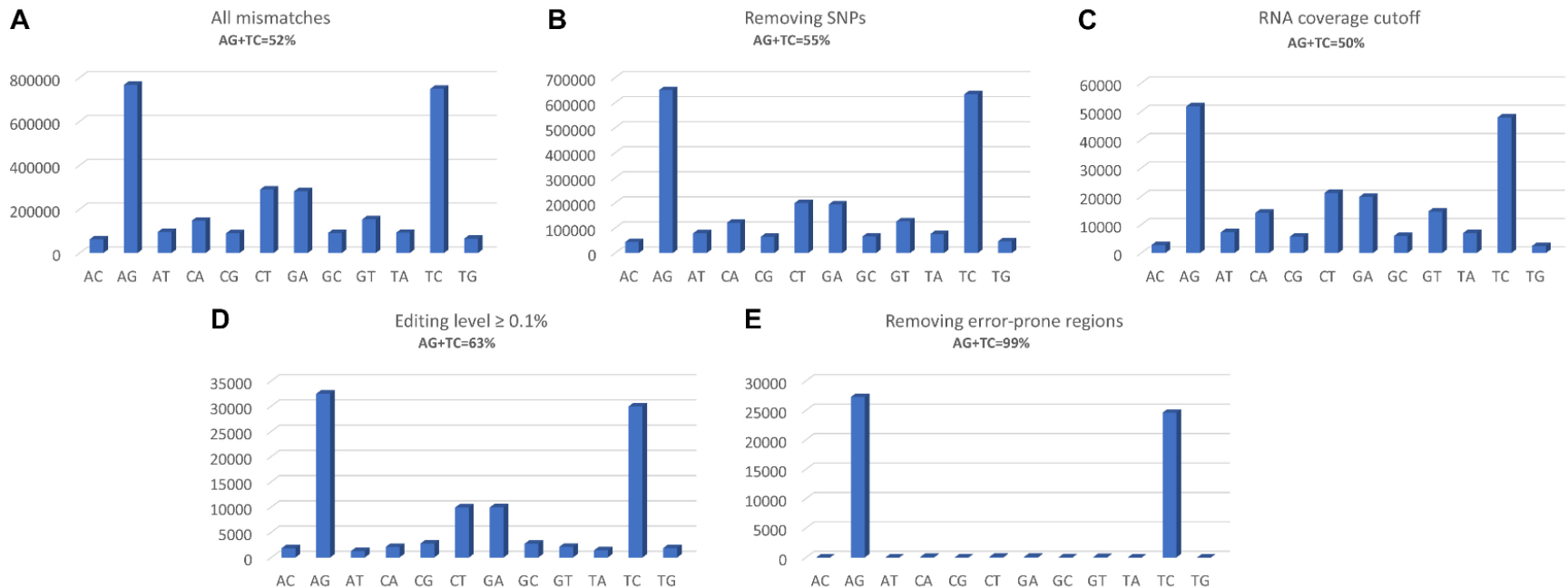


Distributions of RNA variants detected by REDItools obtained following the different filtering steps: **(A)** all mismatches found following mapping, with a phred quality score of at least 30; **(B)** selecting only sites supported by at least 10 WGS reads and removing positions in dbSNP; **(C)** selecting sites covered by at least 10 reads and not falling in homopolymeric stretches longer than five residues or in the first and last six bases of a read; **(D)** selecting sites with an editing frequency of at least 0.1; **(E)** excluding sites in mis-mapped reads (by Blat correction) or near splice sites or in genomic regions containing poorly aligned reads.

Lo Giudice et al. 2020

# RNA editing in human

To profile RNA editing in human tissues we sequenced total RNA from 6 tissues in 3 Caucasian and *"non diseased"* Individuals (sex and age matched) using the Illumina HiSeq2500 platform.
Paired end RNA-Seq reads (2x100) were generated according to the strand-oriented Illumina TruSeq kit. In addition, we produced WES and WGS (20x) reads from the same samples.

| ID | TISSUE | READ PAIRS | PERCENT_DUPLICATION | PCT_RIBOSOMAL_BASES | PCT_MRNA_BASES | PCT_CORRECT_STRAND_READS |
|---|---|---|---|---|---|---|
| 11 | brain | 73375354 | 45,48 | 0,000002 | 0,792654 | 0,9932 |
| 6 | brain | 109695738 | 15,93 | 0,000004 | 0,62469 | 0,990539 |
| 7 | brain | 98295466 | 26,26 | 0,000003 | 0,668658 | 0,988722 |
| 16 | heart | 87631200 | 25,4 | 0,000008 | 0,781802 | 0,994023 |
| 17 | heart | 74406800 | 23,62 | 0,000017 | 0,644371 | 0,993492 |
| 18 | heart | 88158897 | 23,88 | 0,000008 | 0,657386 | 0,992265 |
| 3 | kidney | 77064350 | 30,23 | 0,000003 | 0,761645 | 0,992948 |
| 4 | kidney | 87149042 | 19,93 | 0,000004 | 0,715985 | 0,992623 |
| 5 | kidney | 76433530 | 36,18 | 0,000002 | 0,731433 | 0,992698 |
| 1 | liver | 78455550 | 58,55 | 0,000001 | 0,890288 | 0,997211 |
| 10 | liver | 74220297 | 48,53 | 0,000004 | 0,646801 | 0,993038 |
| 2 | liver | 81843515 | 43,15 | 0,000002 | 0,746217 | 0,995784 |
| 12 | lung | 72506728 | 37,8 | 0,000004 | 0,622289 | 0,991793 |
| 8 | lung | 74192504 | 40,93 | 0,000004 | 0,750165 | 0,995195 |
| 9 | lung | 75368198 | 39,53 | 0,000003 | 0,73011 | 0,994374 |
| 13 | muscle | 80371298 | 49,57 | 0,000002 | 0,910309 | 0,994056 |
| 14 | muscle | 88521121 | 39,78 | 0,000006 | 0,877469 | 0,995332 |
| 15 | muscle | 105596745 | 43,19 | 0,000006 | 0,944626 | 0,995823 |

Statistics obtained by Picard on GencodeV19

# RNA editing in human

Most of the detected RNA editing events were A-to-G (>97%). Potential non canonical events were rare and showed frequency values less than 0.05. in addition, the fraction of A-to-G changes in non-synonymous sites was notably high.

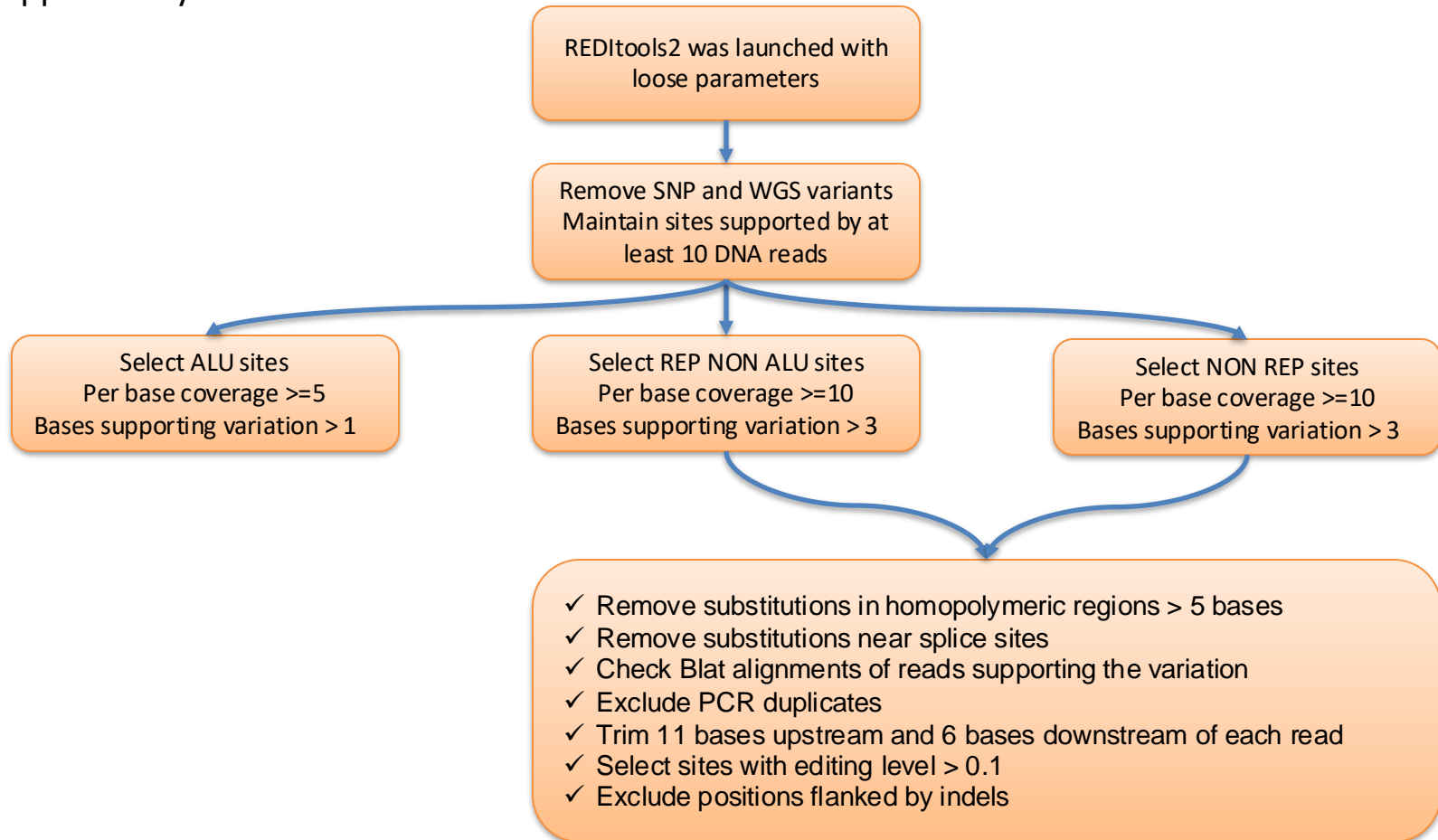Picardi et al. 2015 Sci. Rep.
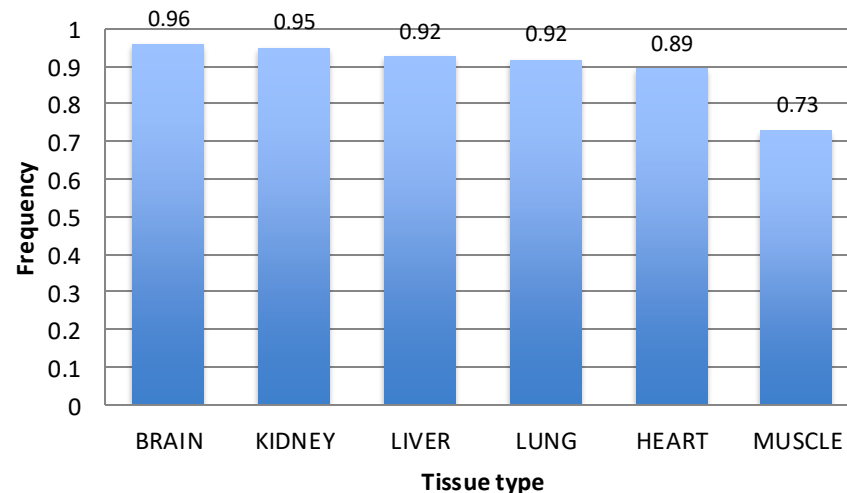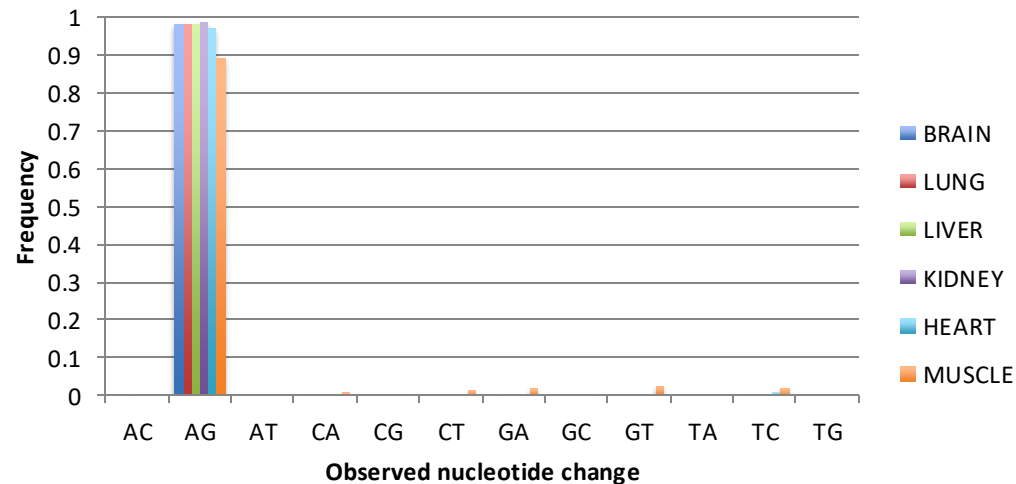
# The Human Inosinome

We investigated the inosinome similarity across human tissues. Cluster analysis based on pairwise comparison of RNA editing levels per sample by the Spearman correlation coefficient, showed well-defined tissue segregation.



Picardi et al. 2015 Sci. Rep.
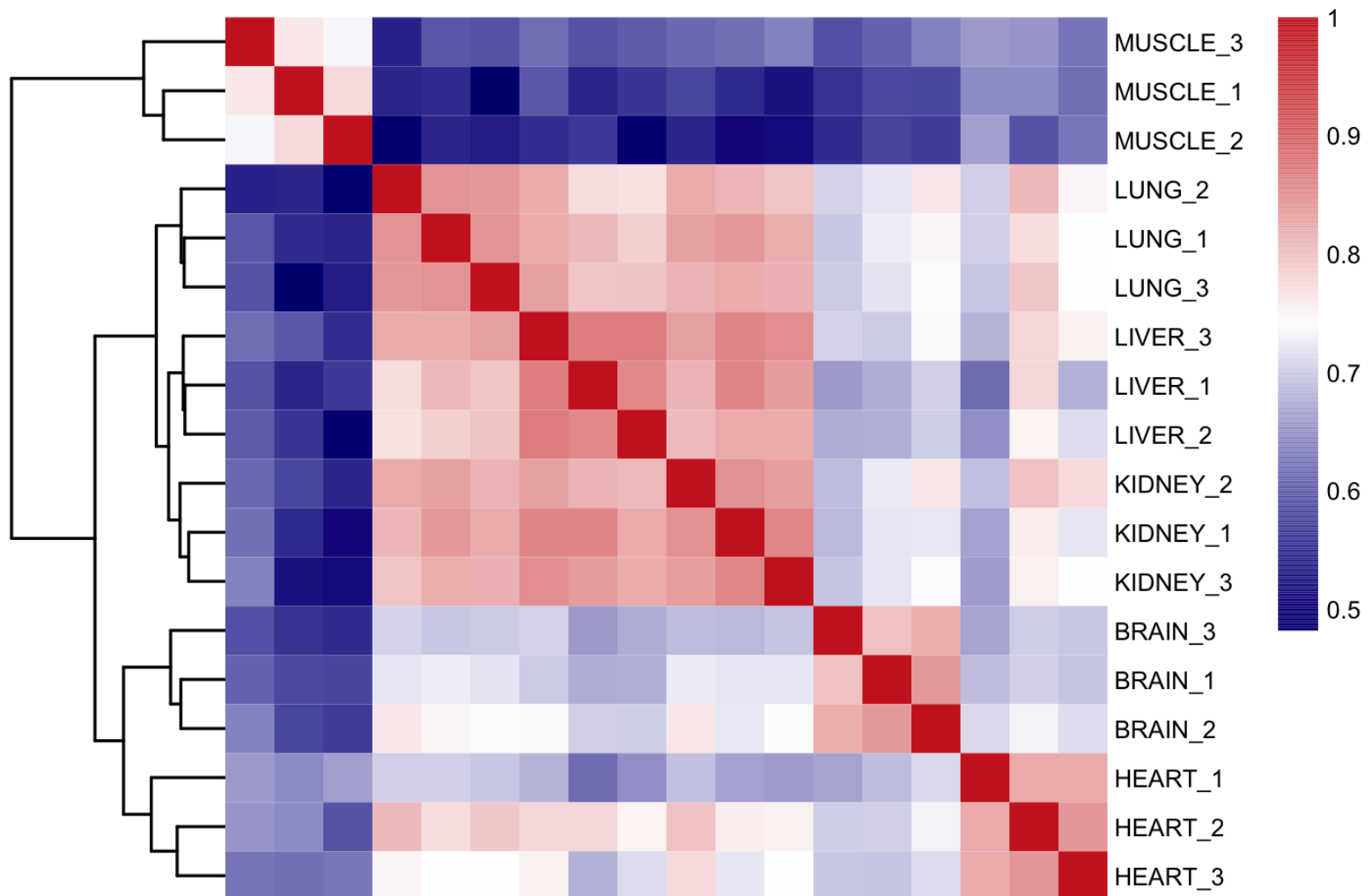
# Is the human inosinome complete?
## RNA editing in GTEx samples

Using REDItools2 we have profiled RNA editing in **9643** GTEx samples (v7). Of these **8523** were supported by WGS data.

REDItools2 was launched with loose parameters

↓

Remove SNP and WGS variants
Maintain sites supported by at least 10 DNA reads

Select ALU sites
Per base coverage >=5
Bases supporting variation > 1

Select REP NON ALU sites
Per base coverage >=10
Bases supporting variation > 3

Select NON REP sites
Per base coverage >=10
Bases supporting variation > 3

✓ Remove substitutions in homopolymeric regions > 5 bases
✓ Remove substitutions near splice sites
✓ Check Blat alignments of reads supporting the variation
✓ Exclude PCR duplicates
✓ Trim 11 bases upstream and 6 bases downstream of each read
✓ Select sites with editing level > 0.1
✓ Exclude positions flanked by indels

We discovered **9,892,506** of editing sites. Of these **2,211,105** are known while **7,681,401** are novel (**4,560,571** present in at least 2 independent samples).

# REDIportal: The Human Inosinome

**http://srv00.recas.ba.infn.it/atlas/**

**REDIportal** is a specialized database for A-to-I RNA editing. In its second release, comprising about **16 M** events, editing sites have been detected using REDItools and a method to identify hyper-edited reads by Porath et al (2014) Nat. Comm.

# Is RNA editing cell type specific?

We have profiled RNA editing in **466** single cells of human brain cortex from **living individuals**, in which a transcriptomic analysis was already been completed (Darmanis et al. PNAS 2015). To provide a more realistic estimate of global editing activity per cell, we calculated the Alu editing index (**AEI**) per cell as it represents the weighted average editing level across all expressed Alu sequences. To confirm cell specificity of RNA editing, we performed a non-metric multidimensional scaling (nMDS) analysis, revealing four clusters corresponding to astrocytes, neurons, oligodendrocytes and OPCs.



Picardi et al. RNA 2017
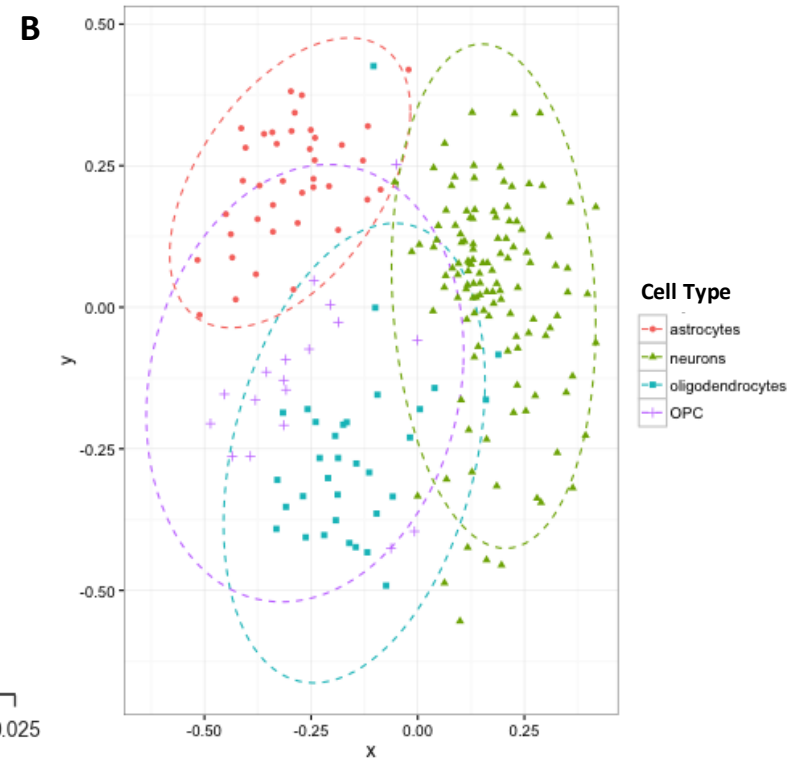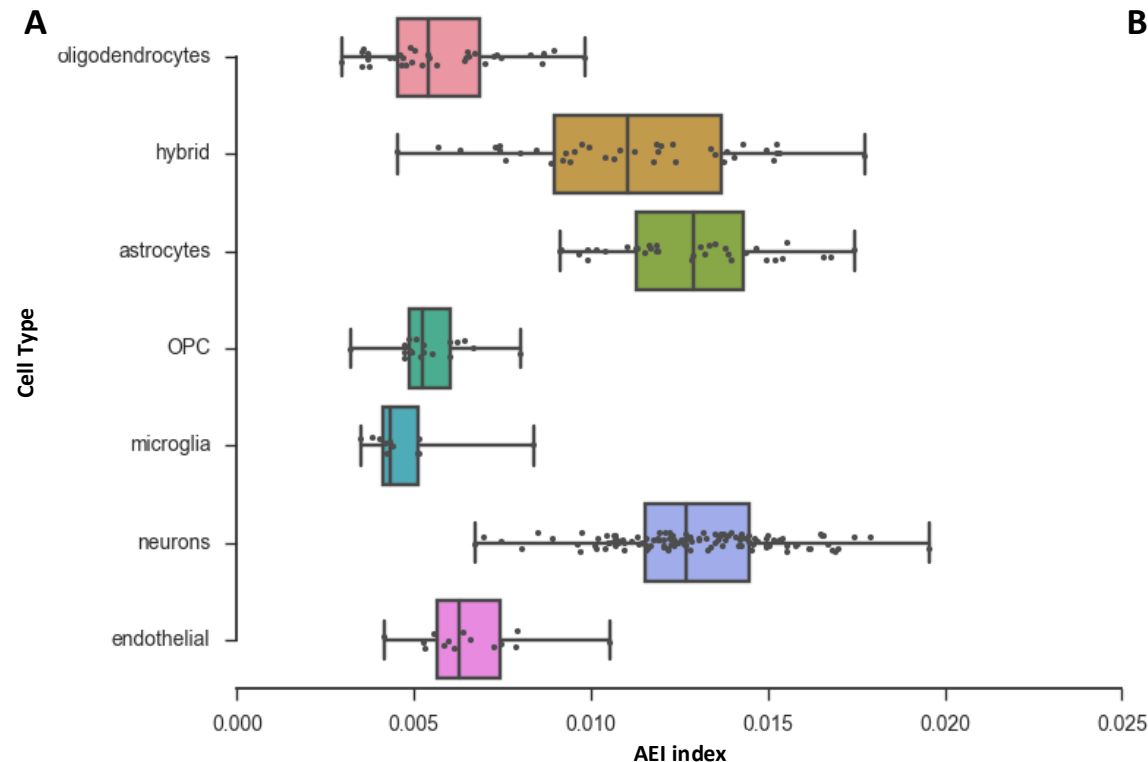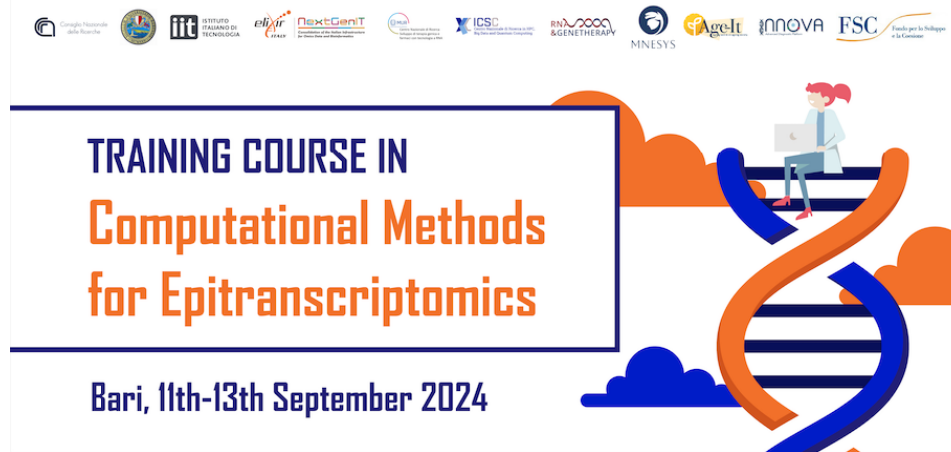
**RNA Editing Agenda**:

*15.30-16.30* **Detection of rare C-to U RNA editing events – Adriano Fonzino**

*16.30-17.00 Coffee Break*

*17.00-18.00* **RNA editing detection by REDItools – Alessandro Silvestris**

*Day 3: Friday 13th September*

9.00-9.30 **Using REDItools and REDIportal to profile RNA editing – Alessandro Silvestris**

9.30-11.00 ***De novo* detection of RNA editing sites by RNAseq data – Silvestris/Fonzino**

11.00-11.30 *Coffee break*

11.30-13.00 **Profiling known RNA editing events in RNAseq experiments – Silvestris/Fonzino**