Regulation and biological functions of alternative splicing in neurons of the adult mice visual cortex

Centre for Developmental Neurobiology, King's College London, New Hunt's House, SE1 1UL London, UK
Jude Popham, Fursham Hamid, Eugene Makeyev

Introduction

Brain function depends on coordination of multiple types of neurons which emerge from a smaller number of progenitor states. The nervous system expresses an especially large collection of RNA isoforms from a single gene by alternative splicing, but how this molecular program contributes to and regulates the emergence of individual neuronal identities via processes like NMD remains poorly understood. This project addressed this by analysing high throughput single cell RNA seq (scRNA-seq) from the mouse primary visual cortex (VISp) by Bakken et al., in 2018 [1]. The expression and splicing profiles of selected neuron subclasses were quantified and compared between cortical layers, subclasses and cellular compartments. Raw data included cell exon counts (**cec**), cell intron counts (**cic**), nuclei exon counts (**nec**) and nuclei intron counts (**nic**). We found that the project supports the mechanism and presence of NMD containing transcripts in the nuclei of PV interneurons and maps the different isoform profiles of GABAergic and glutamatergic layers.

Method

Data Processing + Gene Expression Analysis by Subclass Layer

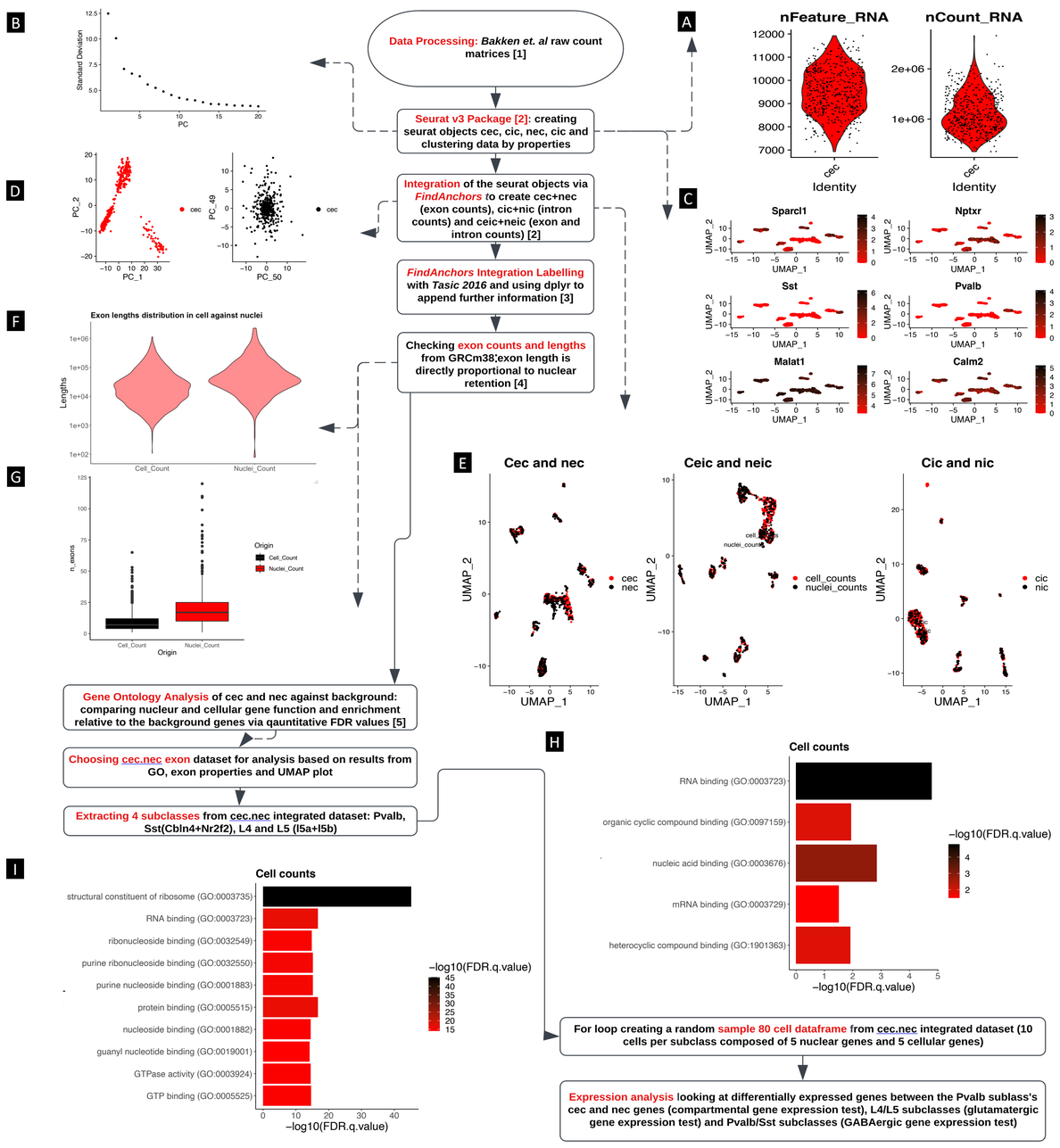


Figure 1: Data Processing Flowchart: A: Violin Plot showing the distribution of the features (genes) and their count frequency in cells. **B: Elbow Plot** which shows PC standard deviations **C: Feature Plot** of the top 2 most differentially expressed features in *L4/L5(top), Pvalb/Sst (middle)* and *Cec/Nec Pvalb (bottom).* **D: PCA** plot showing clustering parameters by relationship of cell properties in the data; key for clustering data, often by subclasses **E: UMAP** plot used to compare mixing and integration of datasets **F: Violin Plot** of exon lengths in cell counts compared to nuclei counts. **G: Box Plot** showing exons distribution in cell against nuclei. **H: Gene ontology** of nuclei counts. **I: Gene Ontology** of cell counts.

Splicing Analysis of Selected Subclass Layers

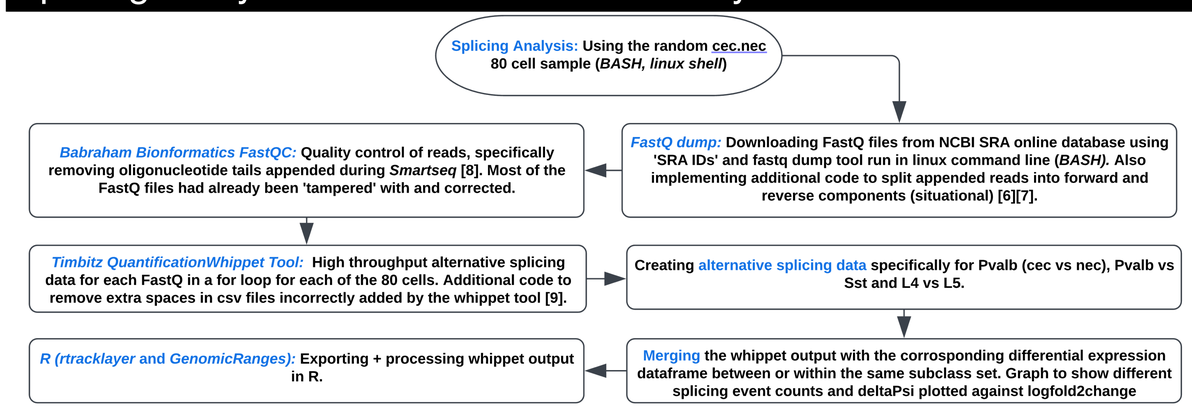
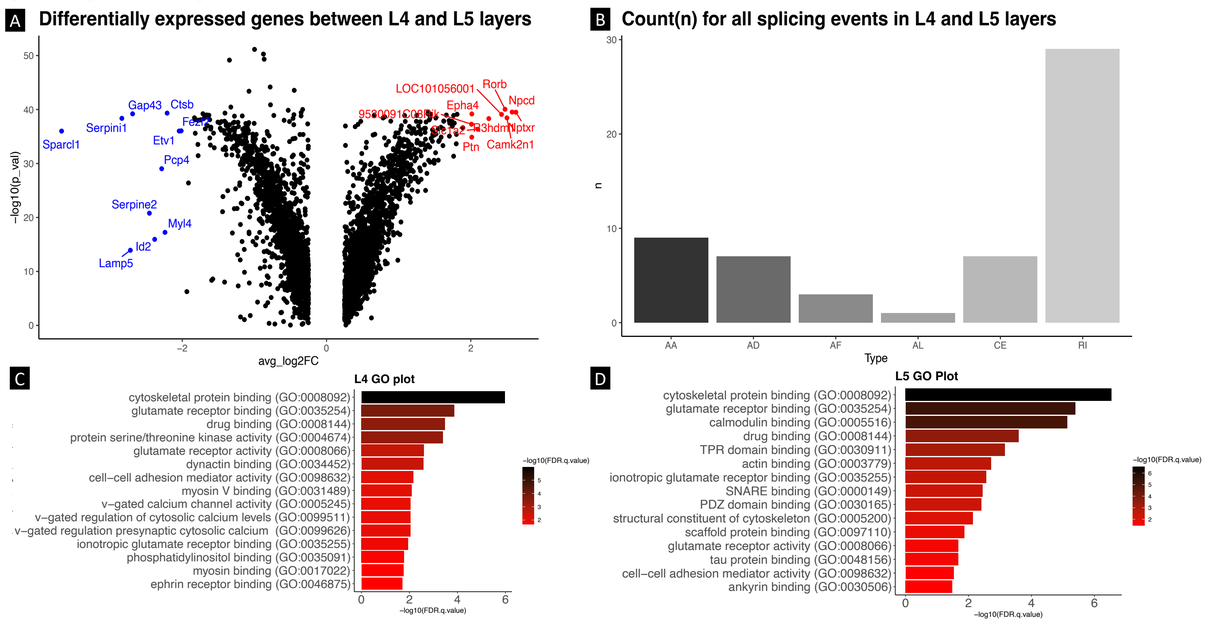


Figure 2: Splicing Analysis Flow Chart: Various online tools were used with scripts run in Linux shell via the Bash command to produce data on splicing events and their frequencies in different subclasses.

Results

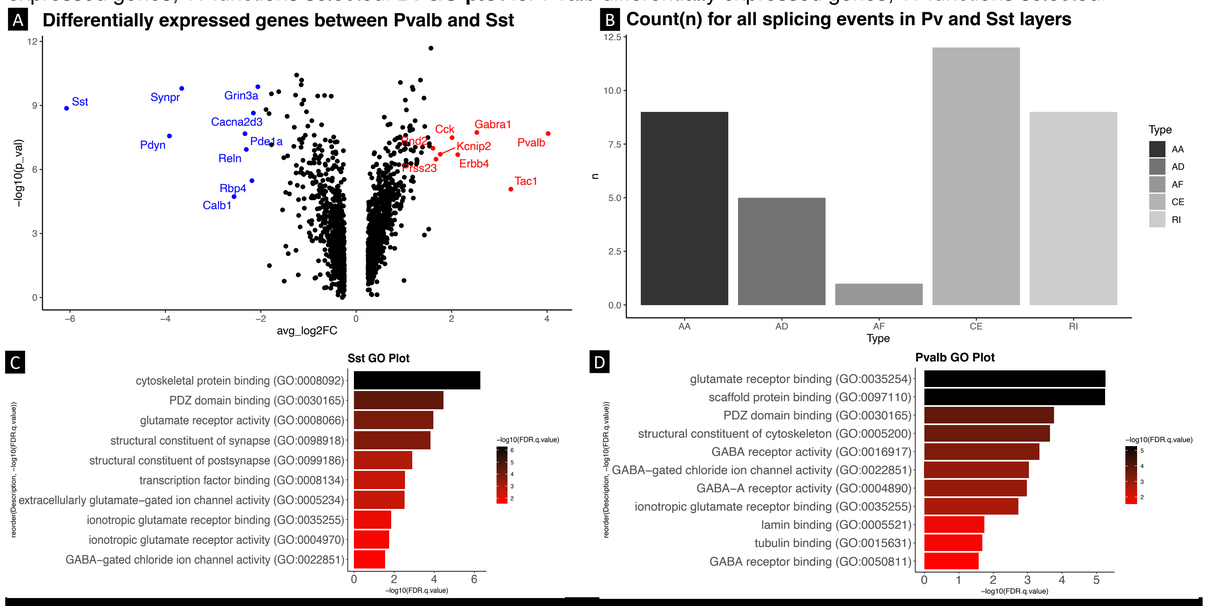
L4 Subclass Compared to L5 Subclass (Glutamatergic)

Figure 3: L4 vs L5 comparison (cell exon counts) A: Volcano Plot of differential expression of genes between the L4 and L5 subclasses. High values of *avg_log2FC* correspond to a higher expression in L4. *Avg_log2FC* is fold change of expression. It is standardized by log2 as most expression profiles are two-fold changes (not ten-fold in which log10 would be used). **B: Box Plot** showing all alternative splicing events and their frequencies (n) **C: Gene ontology** for L4 differentially expressed genes against background: 15 selected functions of interest. **D: Gene Ontology** for L5 differentially expressed genes against background: 15 selected functions of interest.



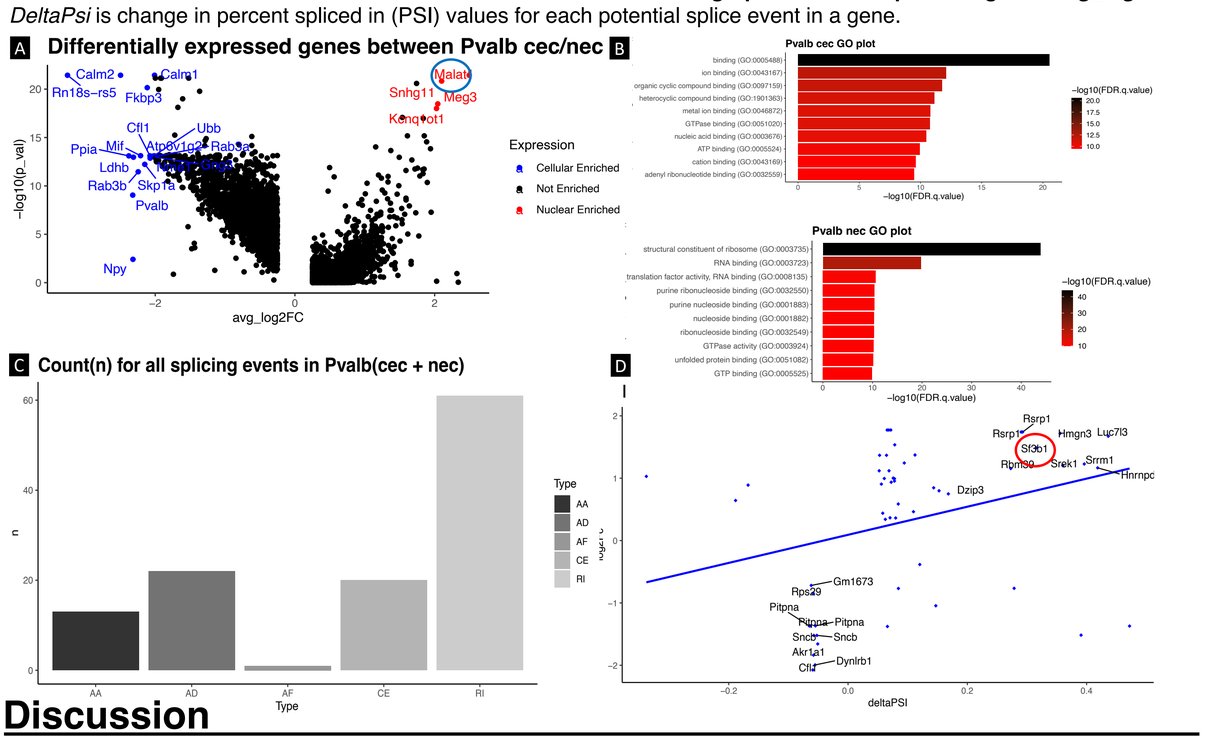
Pvalb Subclass Compared to Sst Subclass (GABAergic)

Figure 4: Pvalb vs Sst comparison (cell exon counts) A: Volcano plot of differential expression between the Sst and Pvalb subclasses. High values of *avg_log2FC* correspond to a higher expression in Pvalb. **B: Box Plot** of all alternative splicing events and their frequencies (n) within the combined Pvalb and Sst exon counts**. C: GO plot** for Sst differentially expressed genes; 11 functions selected. **D: GO plot** for Pvalb differentially expressed genes; 11 functions selected.



Pvalb Subclass (Cell Counts Compared to Nuclear Counts)

Figure 5: Pvalb nuclear and cellular counts comparison (cell and nuclei exon counts) A: Volcano Plot to show differentially expressed genes between Pvalb's cellular exon counts and nuclear exon counts. High values of *avg_log2FC* correspond to a higher expression in the nucleus of Pvalb cells **B: Gene Ontology Plots** which show the functional ontology of cellular and nuclear differentially expressed genes. **C: Bar Graph** to show frequencies (n) of alternative splicing events in the Pvalb combined nuclear and cellular exon counts dataset. **D: Scatter graph of *DeltaPsi* plotted against *avg_log2FC*.** *DeltaPsi* is change in percent spliced in (PSI) values for each potential splice event in a gene.



Discussion

What is the purpose of this data analysis?

- In context, scRNAseq and snRNAseq are two very expensive sequencing techniques that provide a high resolution [1].
- The preferred sequencing technique is batch RNAseq of an entire transcriptome to represent a pool of cells. We utilised new high resolution scRNAseq data to analyse gene ontologies that deliver unique **biological functions** in different cortical layers and study changes in gene **regulation** between cortical layers and cellular compartments (particularly potentially NMD regulated genes: high levels of splicing (*deltaPSI*) and high expression (*avg_log2FC*) in the nucleus).

How did we choose our data?

- The choice to use exon counts for this analysis stemmed from the seamless mixing (with little batch variation) between cell and nuclear exon counts; evident from the UMAP plot, **fig1E**. The concentration of longer exons in the nucleus (**fig1F**) also made sense biologically as they are retained due to longer exportation time and a longer alternative splicing event.
- Furthermore, when comparing functional gene ontologies between cell and nuclear genes there was a clear difference in enriched gene functionality; for example *mRNA binding* was far more enriched in the nucleus, **fig1H**.
- A random 80 cell sample which represented 4 cortical layers and both cell/nuclear counts was used for splicing analysis.

What can we gather from differences in L4 and L5 layer cell subclasses (glutamatergic, fig3C and fig3D)?

- There is evidently no expression of GABAergic genes in glutamatergic neurons. The L5 layer is closer to the basal lamina of the mouse brain. It projects further across the brain than L4, perhaps facilitated by *scaffold protein binding* enrichment.
- The L4 has a much higher cortical depth and thus a higher excitation density [13]. This is supported by the higher enrichment of *glutamate receptor activity* of the L4 layer.
- L4 neuron transmission is heavily regulated by calcium levels. Both layers are strongly connected to the cytoskeleton of the cortex, facilitated by high expression of genes involved in *cytoskeletal protein binding and cell-cell adhesion.*
- Interestingly, L4 has many *myosin* binding genes which are not present in L5 which is enriched with *actin* binding genes.

What can we gather from differences in Pvalb and Sst cell subclasses (GABAergic, fig4C and fig4D)?

- Pvalb and Sst subclasses are GABAergic interneurons. This means they are spread throughout the cortical layers (L1 to L6) [12]. They have both *GABAergic and glutamatergic* activity to act as interneurons (between cells), but *GABAergic* activity appears to be more enriched in Pvalb whereas *glutamatergic* activity in Sst (**fig4C/D**).
- Their distribution and anchoring throughout the layers is facilitated by *lamin* and *tubulin* binding genes that are required for their work; anchoring to the cytoskeletal components is also reinforced by *PDZ domains*.
- The way Pvalb and Sst interneurons inhibit a signal depends on the location synapse connectivity of the glutamatergic cells (thus functional differences in their *receptor* gene ontology to target different parts of the glutamatergic neurons).

What can we gather from differences in Pvalb cell and nuclear counts?

- Out of all splicing events, the strongest positive correlation between the splicing percentage (*deltaPSI*) and positive fold nuclear expression (*avg_log2FC*) was represented in cassette exons (CE, **fig6B**). CE alternative splicing events produce the highest proportion of NMD transcripts [12]. Exons which have a high splicing percentage and high nuclear retention (high *avg_log2FC*) are potentially NMD regulated (**fig5D**, top right; e.g **Sf3b1 (circled)**). *Malat1* (**fig5A**, trop right, circled) is a non-coding RNA transcript that can be used as a control for nuclear retention.
- NMD is non-sense mediated decay, instrumental in destroying faulty transcripts and recently found to be effective in regulating gene expression via alternative splicing: **AS-NMD modality** [11]. NMD disposition is facilitated by the presence of a premature stop codon in a transcript which can be random or inserted by alternative splicing (**fig6C, Sf3b1: circled**).

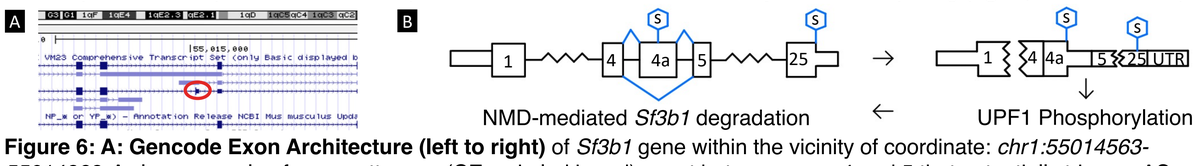


Figure 6: A: Gencode Exon Architecture (left to right) of *Sf3b1* gene within the vicinity of coordinate: *chr1:55014563-55014868*. A clear example of a cassette exon(CE – circled in red) event between exon 4 and 5 that potentially triggers AS-NMD via introduction of a premature stop codon. **B: Drawn Model of AS-NMD** for *Sf3b1* gene. Breakages are intervening exons not involved in the alternative splicing event. The terminal exon(25) is long due to the 3' UTR after the actual stop codon and thin after the premature stop codon (exon 4A). Stop codons and the splicing event are drawn in blue.

Conclusion & Future Work

This experiment can help conclude that single cell sequencing of cell genomes can be used to identify NMD stable isoforms; a technique yet to be explored extensively due to the high cost of single cell and single nuclear RNAseq compared to batch RNAseq. Notably, a correlation was identified between splicing and expression in some exons - strongest in cassette exons [12]. As the cost of sc/snRNAseq reduces with modern technology, this experiment could be taken further with sequenced and labelled **embryonic** mice brains from Smartseq (Bakken et al. sequenced an adult mouse brain)[1]. This would enable the analysis of developmental mouse genes in neurogenesis via comparative bioinformatics techniques, providing a better picture of NMD's role in regulating genes like transcription factors in development of the brain.

References

[1] Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, et al: Single-nucleus and single-cell transcriptomes compared in matched cortical cell types , 2018 [2] Satija Lab, 2022 [3] Tasic, Bosiljka et al. "Adult mouse cortical cell taxonomy revealed by single cell transcriptomics." Nature neuroscience vol. 19,2, 2016 [4] NCBI Assembly, 2012 [5] GOrilla, 2022 [6] Sequence Read Archive (SRA) NCBI, 2022 [7] Fastqdump Edward's Lab, 2015 [8] Babraham Bionformatics FASTQc, 2022 [9] Timbitz Whippet Function, 2021 [10] Hamid, Fursham M, and Eugene V Makeyev. "Emerging functions of alternative splicing coupled with nonsense-mediated decay." Biochemical Society transactions vol. 42,4, 2014. [11] Lim, K.H., Han, Z., Jeon, H.Y. et al. Antisense oligonucleotide modulation of non-productive alternative splicing upregulates gene expression. Nat Commun 11, 3501 (2020). [12] Cell Type-Specific Structural Organization of the Six Layers in Rat Barrel Cortex [13]