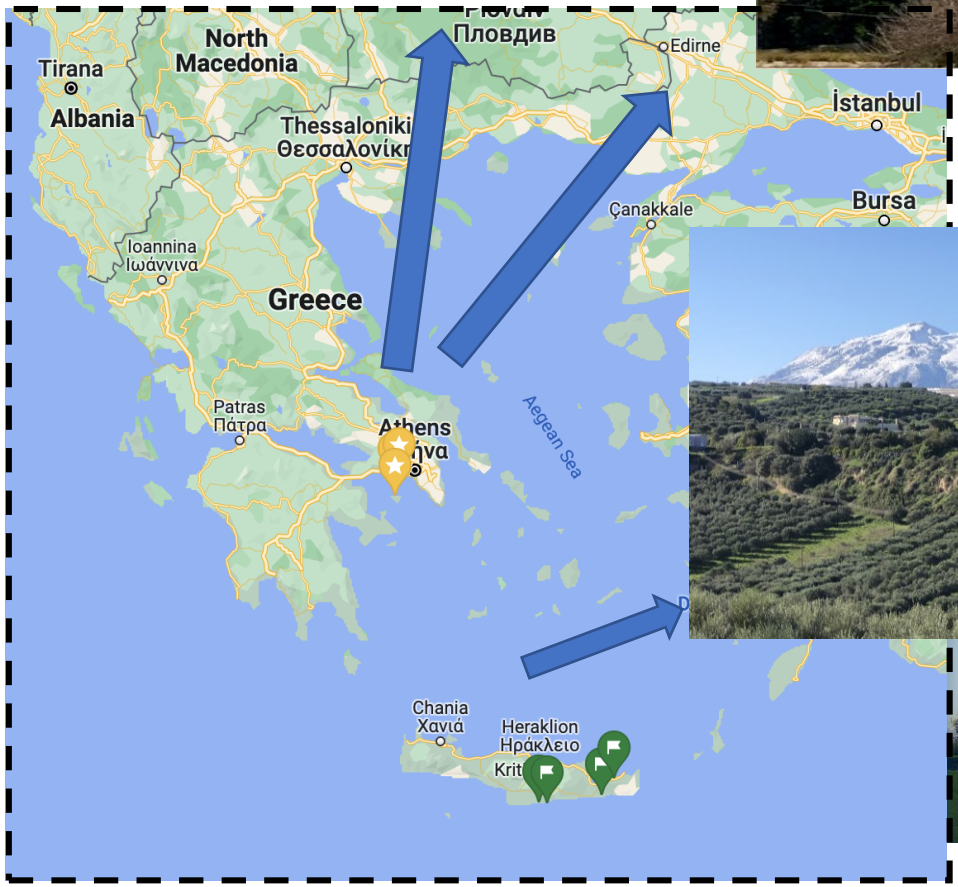
Studying the impact of nascent RNA synthesis on gene regulation

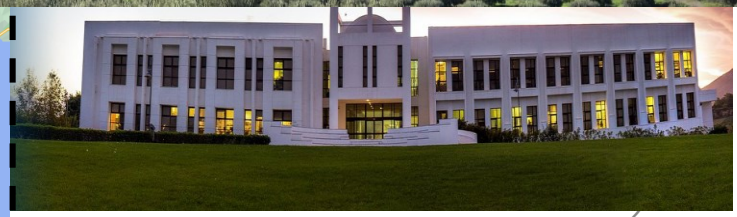
Matthieu D. Lavigne, PhD
Gene Control Mechanisms Lab,
Head of the IMBB Genomics facility,
IMBB-FORTH, Crete, Greece

PhD in Gene medicine, Pr Gorecki lab
University of Portsmouth, UK

MSc Biochemistry
UPMC, Paris, FR

POST-DOCS in IIBEAA (Thanos Lab) and Al. Fleming (Fousteri Lab)





Genomics Facility

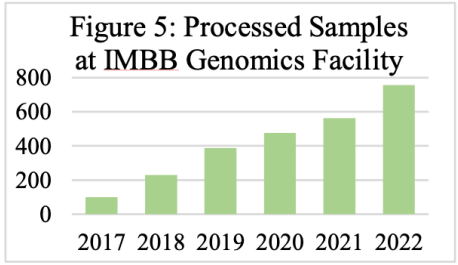
HEAD OF FACILITY

Matthieu Lavigne

FACILITY MEMBERS

Research assistants:
Eirini Stratidaki
Niki Gounalaki

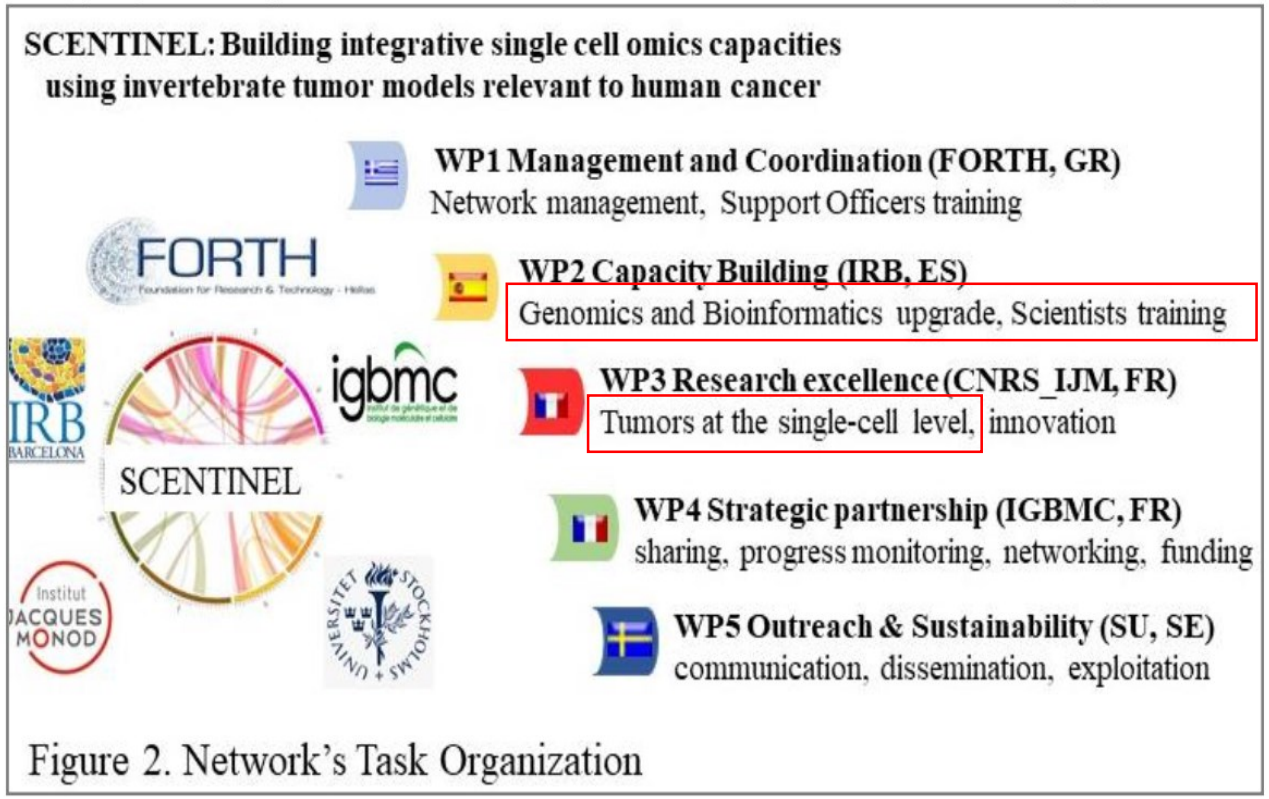
Data management and analysis:
Emmanouil Dialinas



Illumina NextSeq 500 + 2000 (UoC)
Nanopore (ADNA)

-Answers to users needs to look at precious and difficult samples (extremely low input RNA-seq)
-Development of custom workflows (MSAP-seq, SLAM-seq…)
WITH NO CHARGE of technician cost or overheads

TWINNING



Current HIBU (since 2021)
(Horizontal BioInformatics Unit)

Installation and plan for development of High Performance Cluster (HPC)

Elevate IMBB science and innovation capacities and Genomics Facility technological capacity by

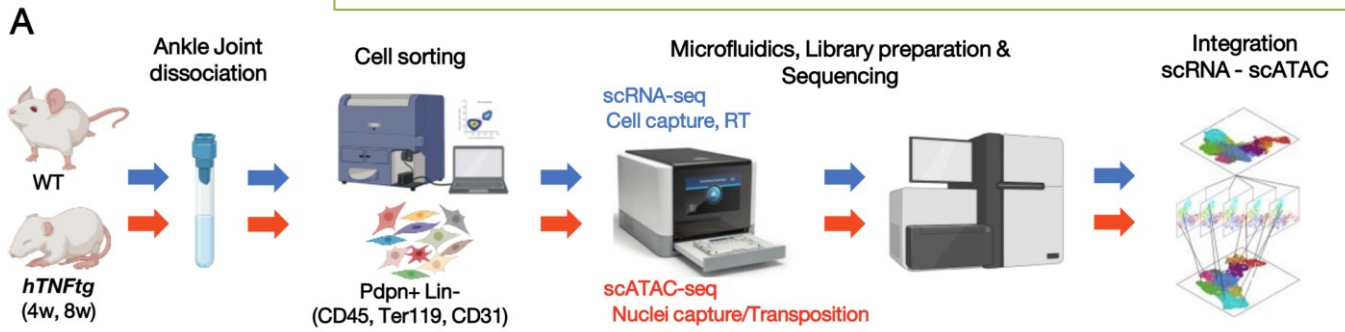
i) bringing expertise to facility's personnel to use new 10 X Genomics platform (purchased by Chamilos/Talianidis)

ii) bring the PIP-seq technology (with A. Pavlopoulos and Enzyquest)

→ **offer single-cell sequencing services to other users** towards the midway of this **twinning** project (2025)

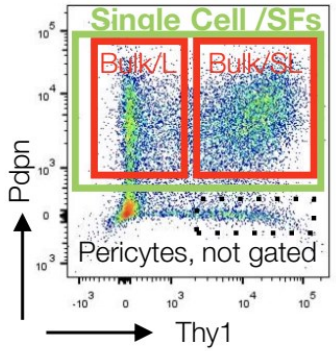
→ Leverage for recruiting one newly trained scientist

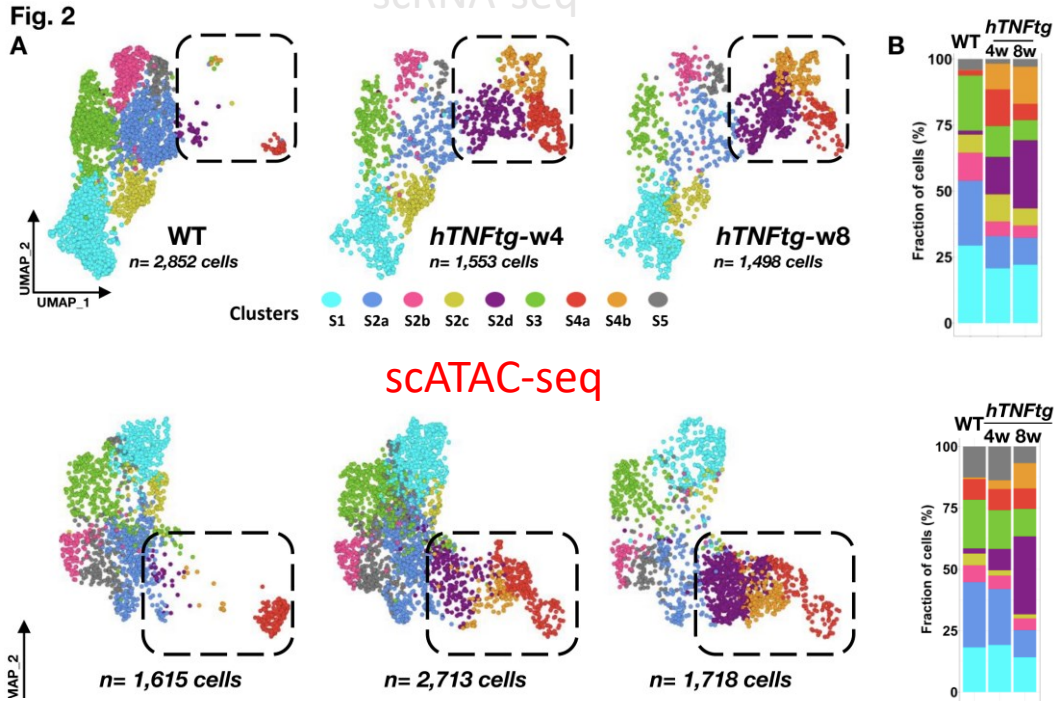
sc multi-omic profiling of Synovial Fibloblasts (SFs)

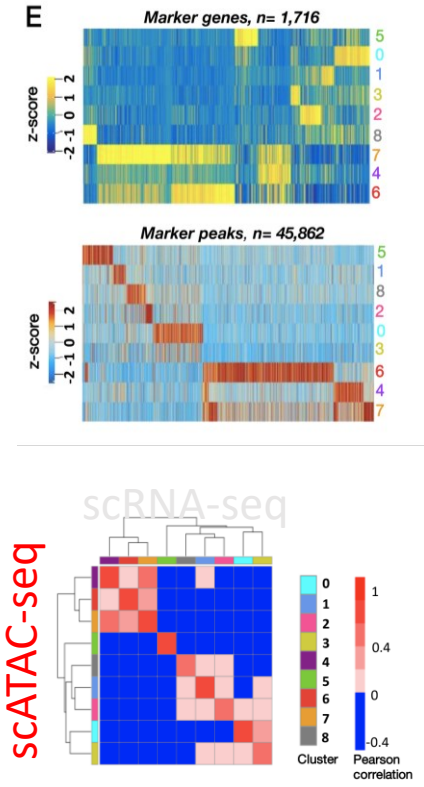


Shared patterns at transcriptional and chromatin levels

Expansion of SFs subtypes in Tg



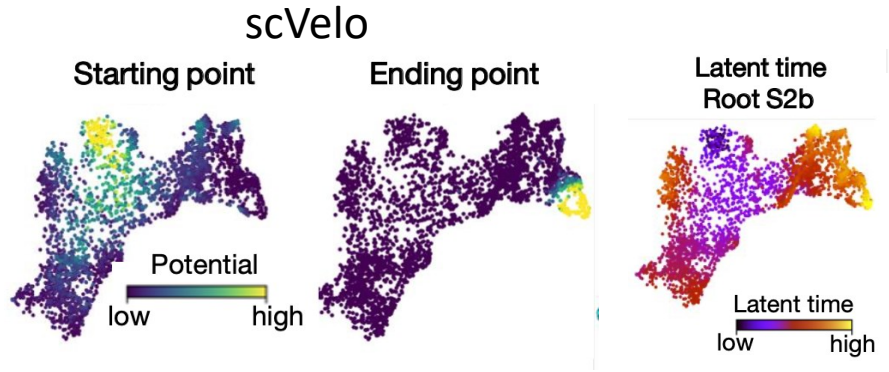




canonical correlation analysis (CCA) enabled the matching of scRNA-seq and scATAC-seq cluster identities

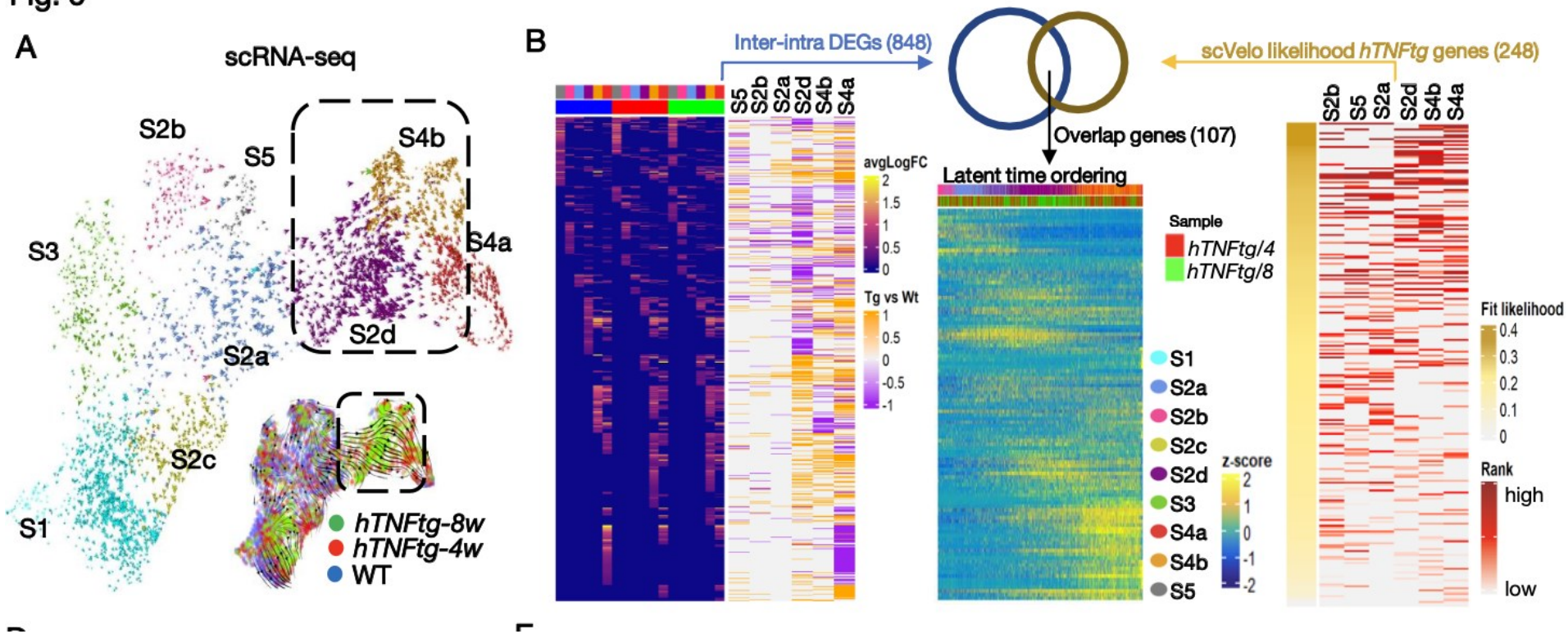
Armaka et al, 2022

A defined trajectory yields pathogenic SFs in diseased joints



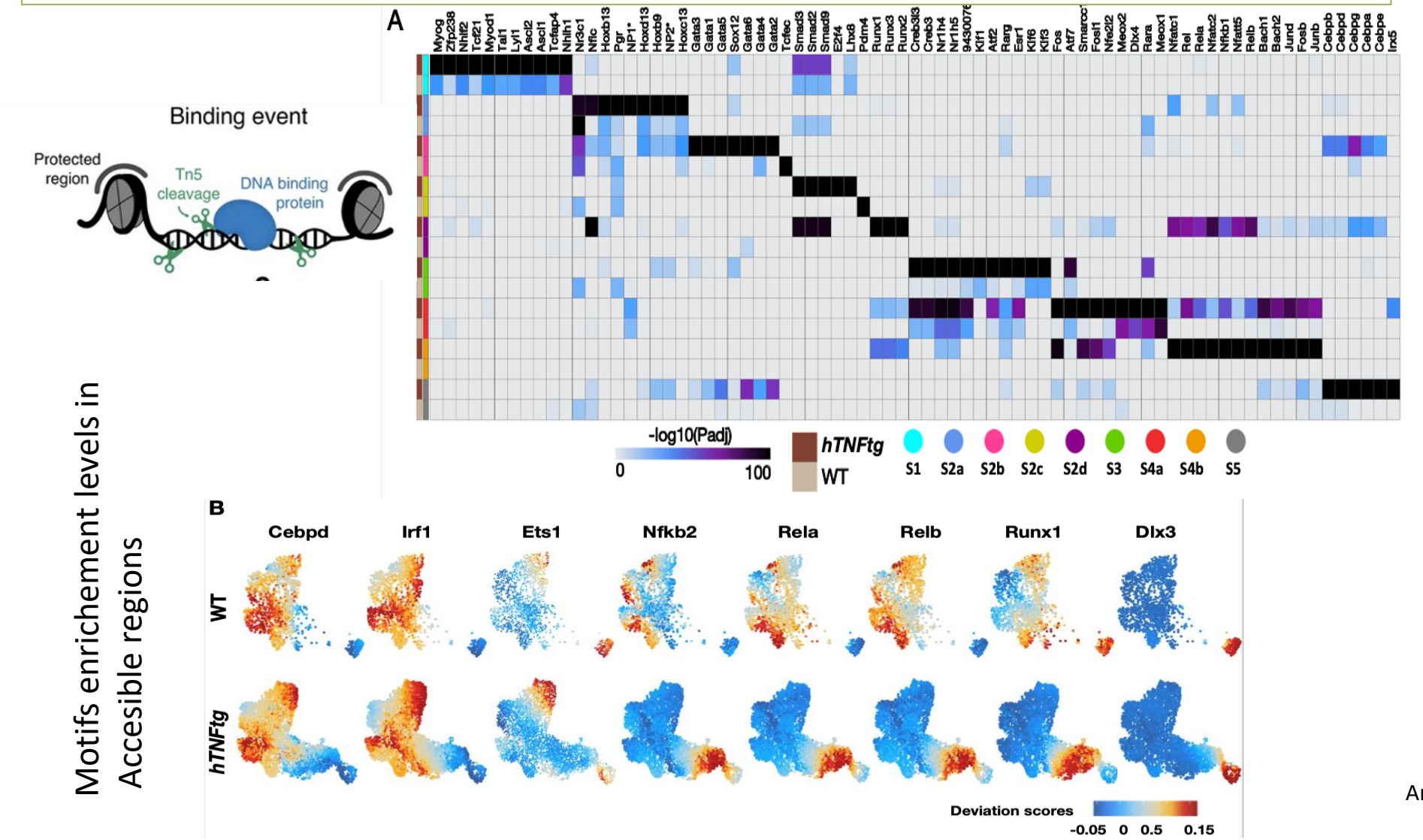
- S2b give rise to the emerging S2d, S4b and S4a SF states in disease
- 107 core genes control the transition to pathological states

Fig. 6



Armaka et al, 2022

DNA motif analyses to determine which TF might control the cell-type or disease-specific regulatory regions and associated genes



Armaka et al, 2022

intermediate and disease lining states are controlled by master regulators including **Runx1**, and Junb/d, **Rel and Nfkb the later** known pro-inflammatory effectors

GENE CONTOL MECHANISM LAB AT IMBB

Established 2021

Hypothesis-driven basic research to find out novel **molecular mechanisms essential for healthy gene expression**

WHY is it IMPORTANT: The **idea to alter activity of chromatin (e.g histone deacetylase (HDAC) inhibitors) to control cancers in the clinic (eg, vorinostat)** has resulted from **prior clear understanding of their mode of action** in transcriptional regulation

Directions of research

How nRNA interactions interactions occur and **what link it to human diseases** such as cancer or neurodegenration

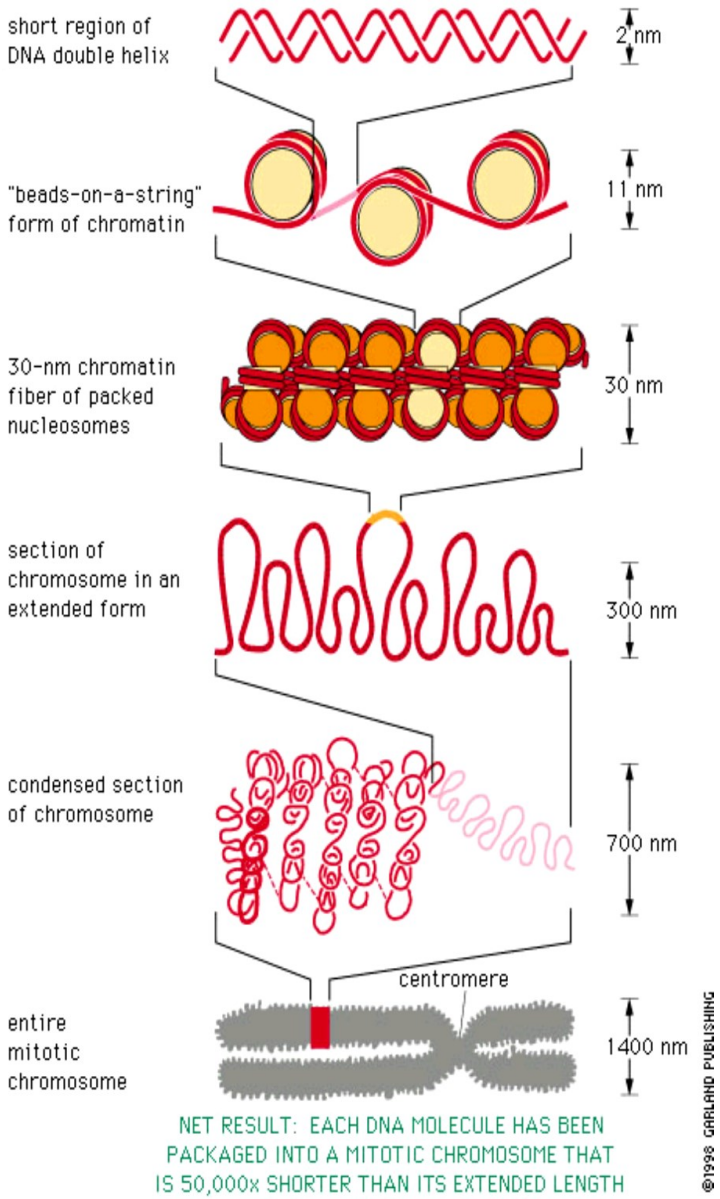
i) Describe **how nascent RNA influence** gene **transcription regulation** and genomic **DNA sequences integrity**

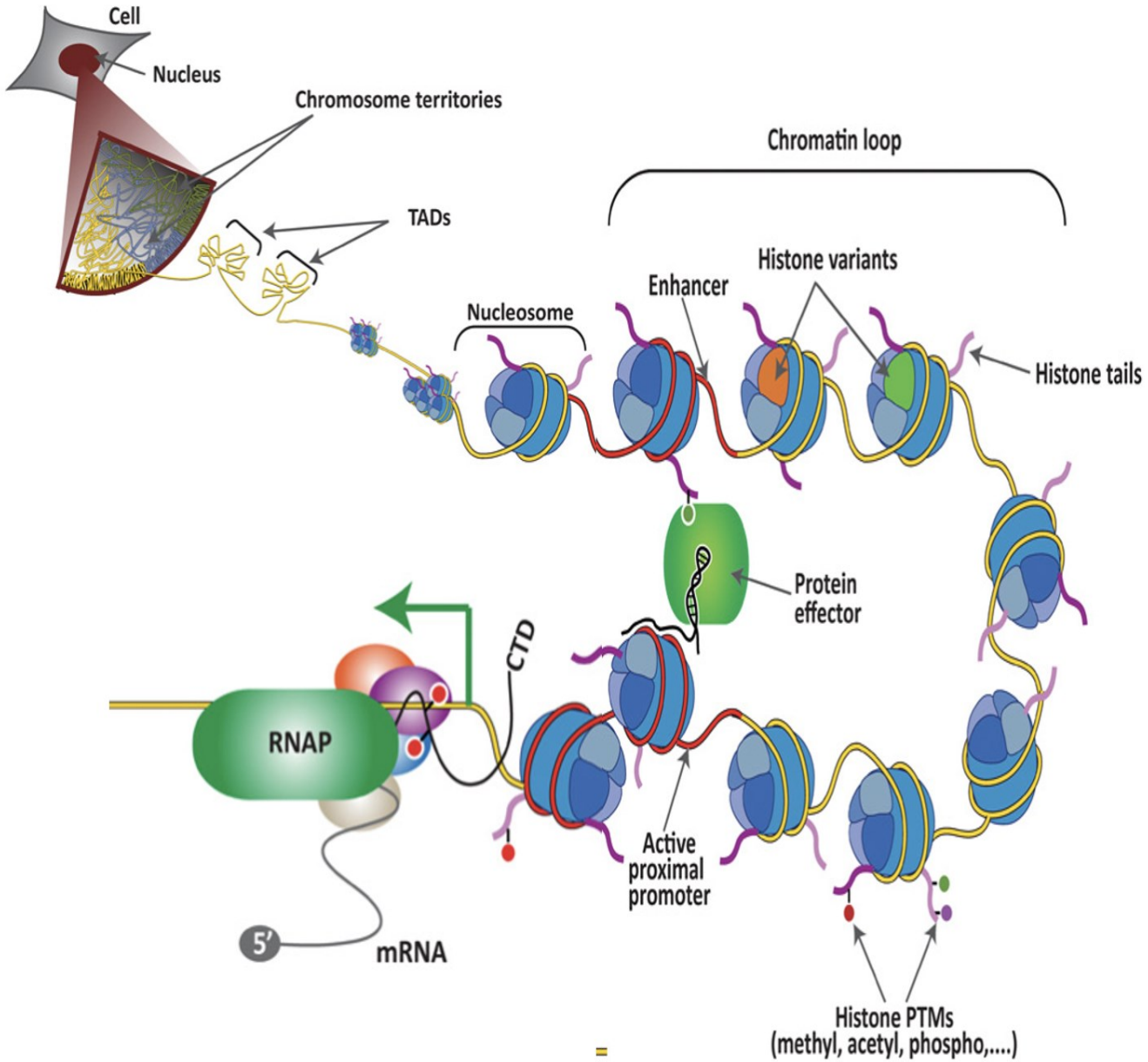
ii) Characterise role of nascent RNA in histone H2B ubiquitination (**H2Bub**) / turn-over rates during transcription

*We **apply both molecular and computational/mathematical biology methods** to integrate data from multi-omics approaches to interrogate protein-DNA-RNA interactions and **learn from data (ML) to model/predict molecular mechanisms** (e.g. ChIP-seq, RIP-seq, HiChIP), RNA expression levels and patterns (mature/nascent RNA-seq, FISH), or chromatin accessibility (ATAC-seq) and modification/conformation status (ChIP-seq, HiC, GRID-seq, Machine learning, Simulation)*

CURRENT/PREVIOUS LAB MEMBERS: Vaios Theodosiou (PhD student), Marianna Stagaki (MSc Bioinformatics, Thesis), Electra Tsaglioti (RA Bioinformatics), Chris Botos (BSc Students), Angeliki Loukopoulou (Msc rotator), Kostis Kydonakis and Myrto Mittleton (MSc MBB, Thesis), Nikos Vouzounerakis, Stergios Manakas, Johnny Petrossian, Electra Kontonikou,

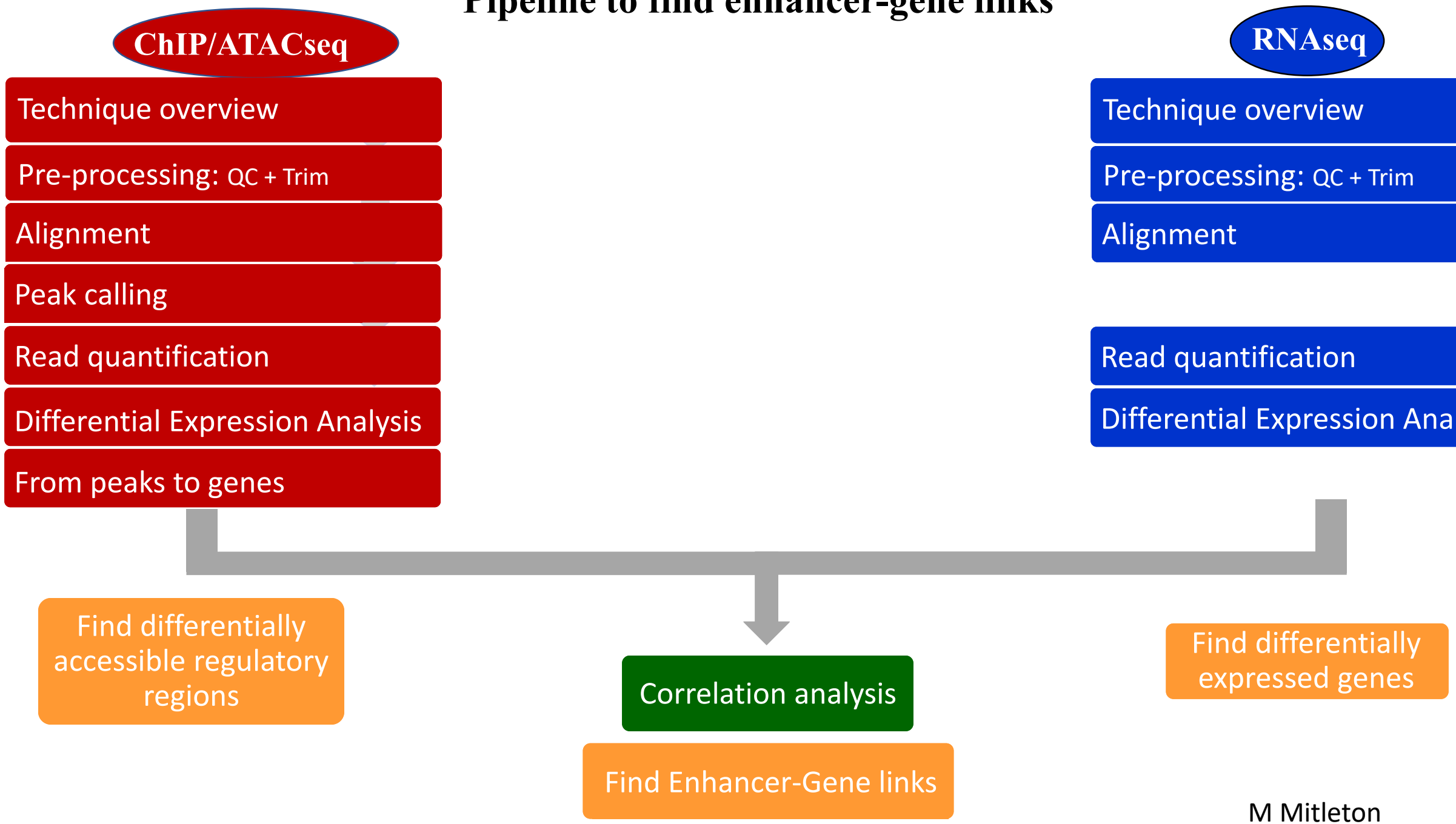
Multi-level 4D genome organisation controls gene expression





TIMELY REGULATED

Pipeline to find enhancer-gene links



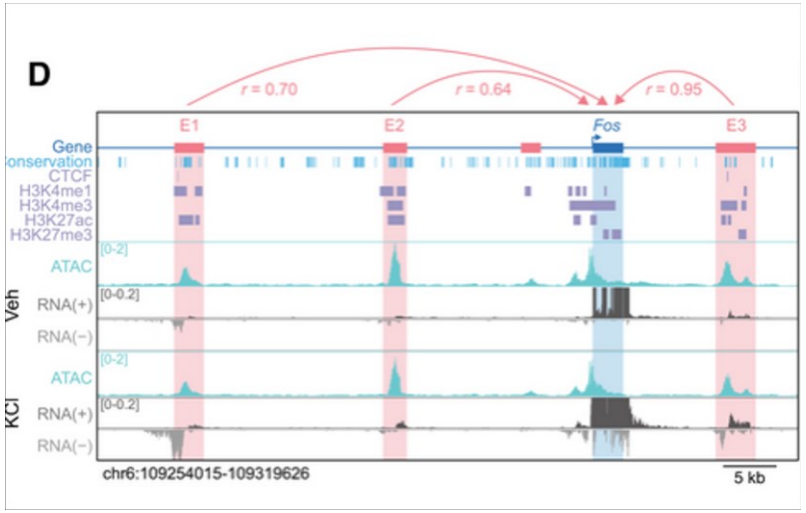
M Mitleton

PROJECTS: INTEGRATION OF RNA-seq and ATAC-seq (BULK or sc) data to discover novel GENE REGULATORY NETWORKS

Enhancer RNAs predict enhancer–gene regulatory links and are critical for enhancer function

Carullo, 2020, NAR

ATAC identifies enhancer regions that can be linked to a given gene expression (**Peak-to-gene** correlation)

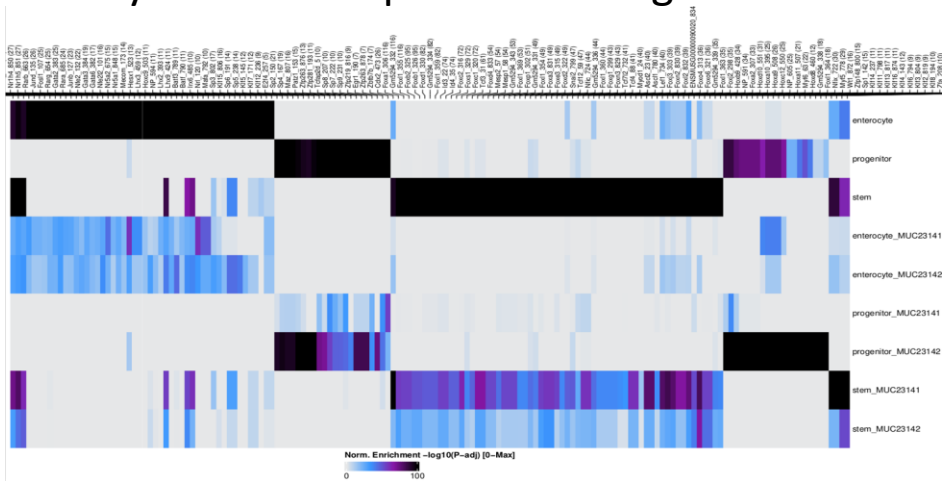


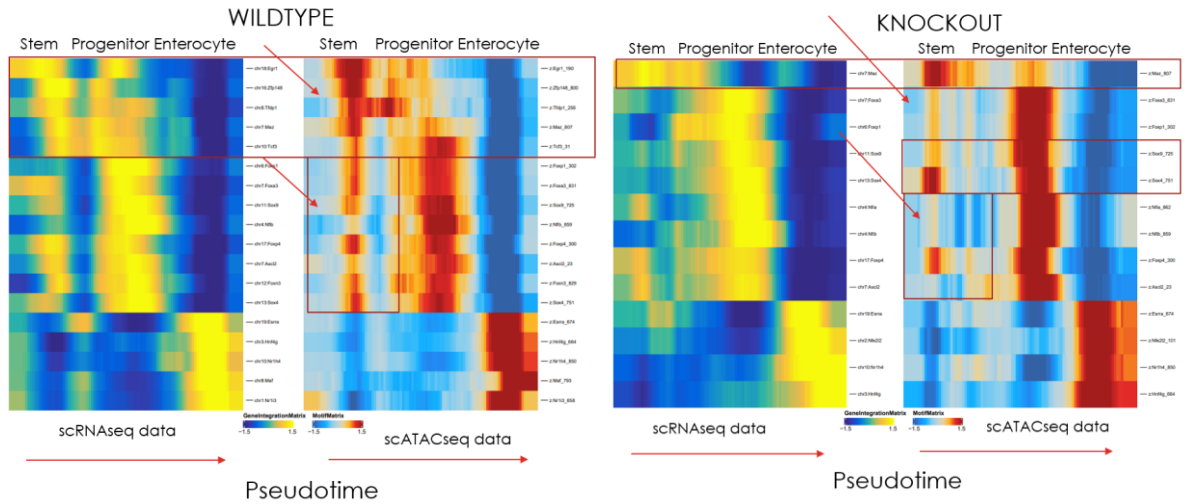
multiomics (R, Python): determine genes regulatory regions

Myrto Mittleton and Johnny Petrosian

sc-omics (R, Python):

TF motifs accessibility and RNA expression changes in WT vs KO mice

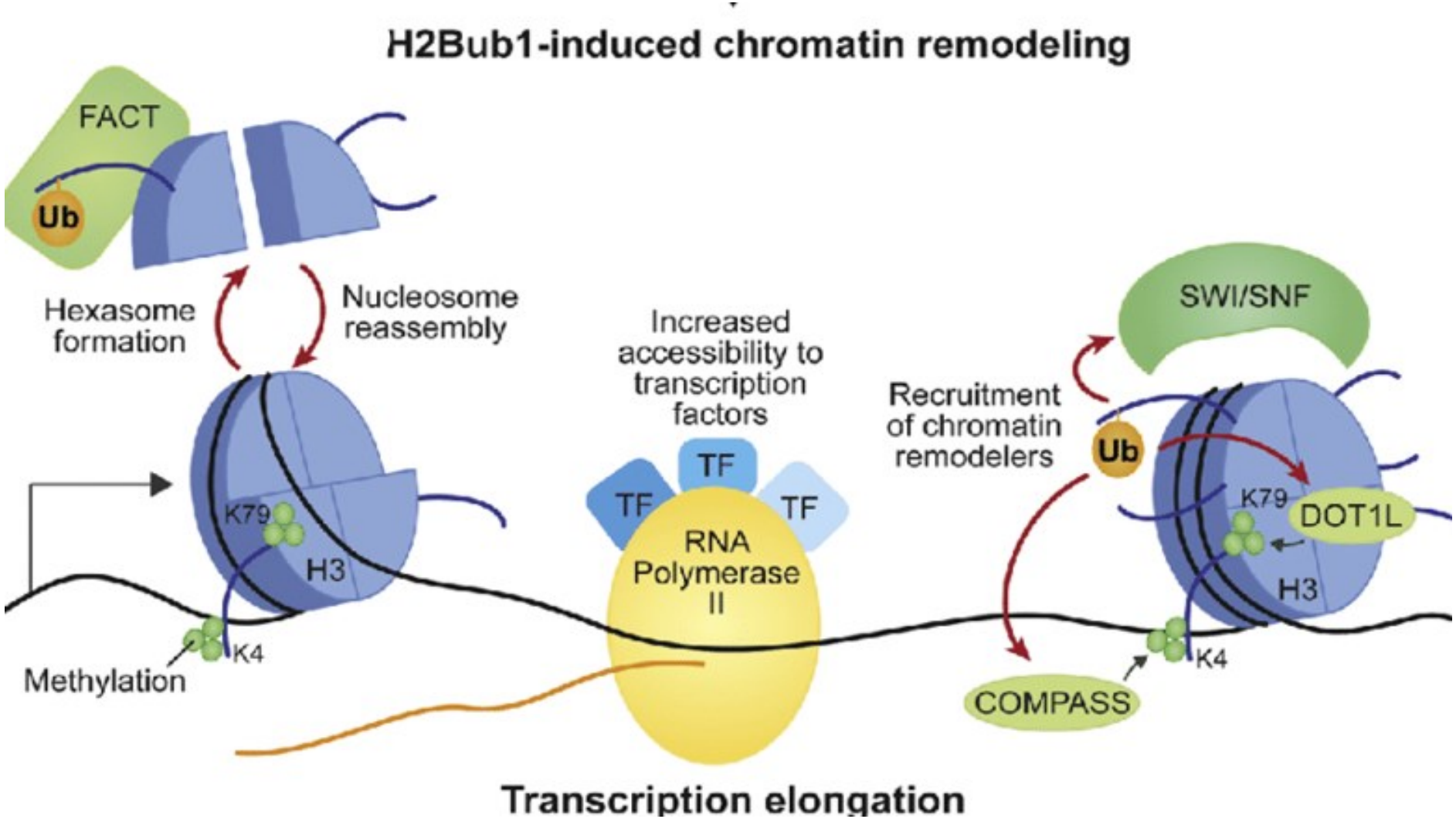




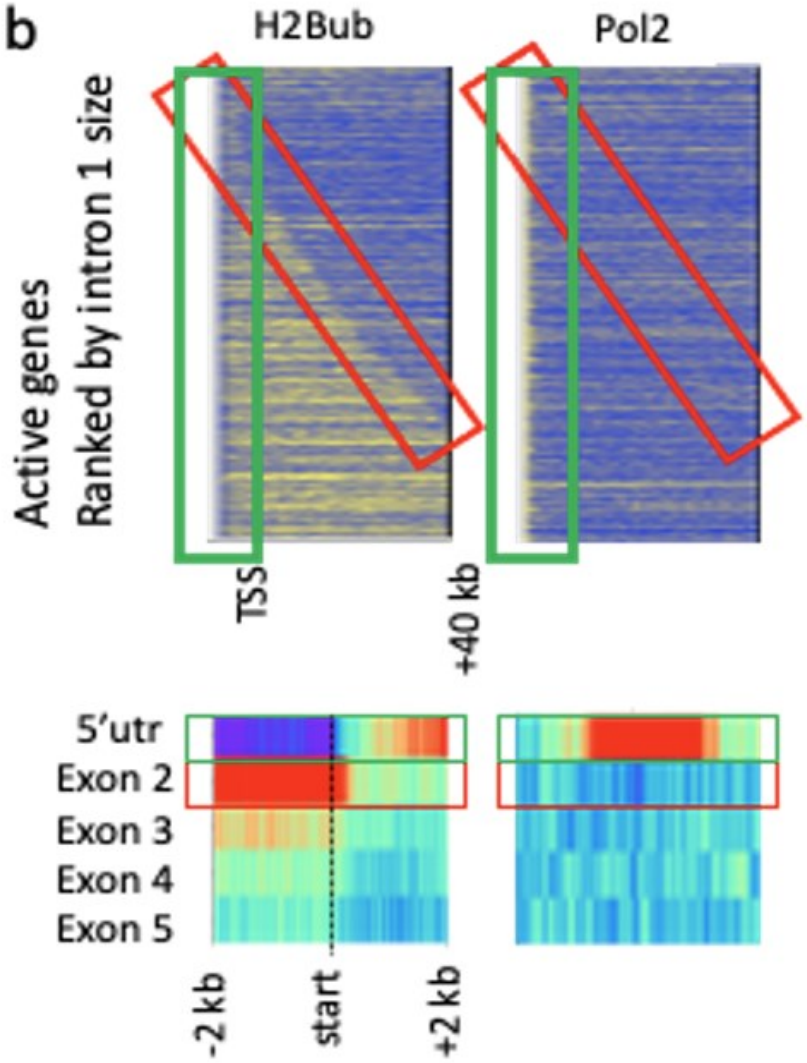
Marianna Stagaki (Collaboration with Talianidis lab)

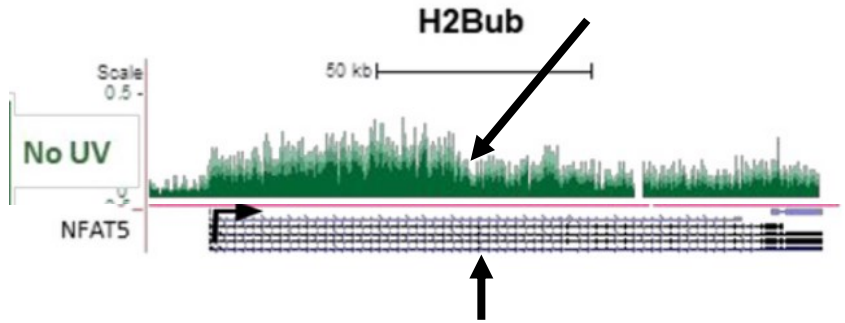
H2Bub and nascent RNA interplay

H2Bub and transcription elongation

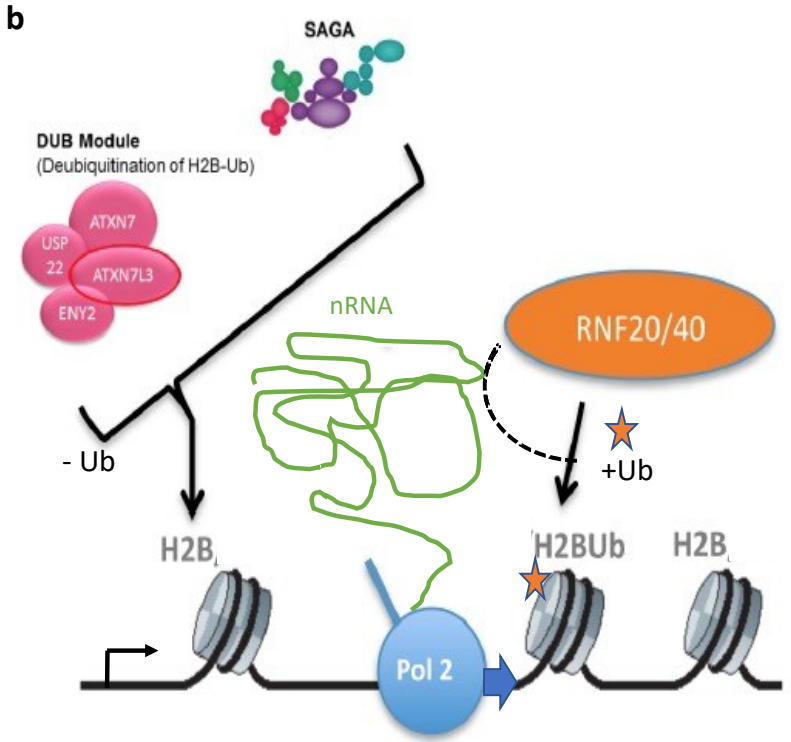


H2Bub Distribution showing interesting topological specificitiy: mechanism of writing/erasing?





First intron-exon junction



Lavigne unpublished and Fanourgakis et al, 2022

Nascent RNA is an integrative component of chromatin

Molecular Cell 2021

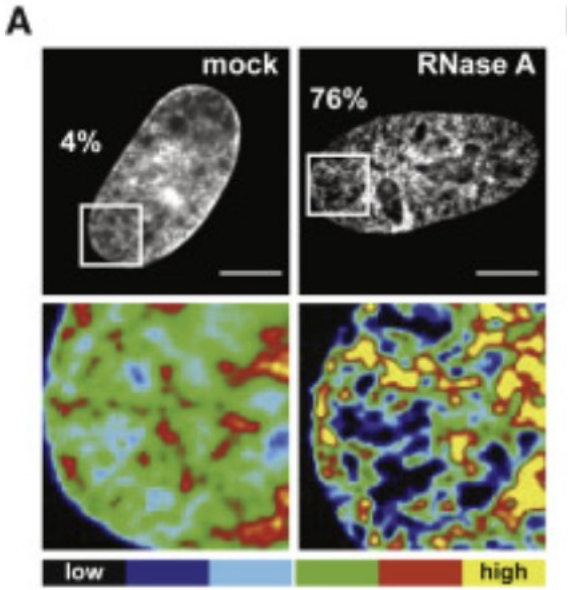
CellPress

Volume 81, Issue 17, 2 September 2021, Pages 3509-3525.e5

Article

Nascent RNA scaffolds contribute to chromosome territory architecture and counter chromatin compaction

Kevin Michael Creamer [1], Heather Jill Kolpa [1], Jeanne Bentley Lawrence [1 2]



nuclear RNA depletion using RNase A disrupts nuclear morphology and causes rapid “collapse” of chromatin into compact regions

PERSPECTIVES

OPINION

Regulatory feedback from nascent RNA to chromatin and transcription

Lenka Skalska, Manuel Beltran-Nebot, Jernej Ule and Richard G. Jenner

modification. Transcription elongation factors bind to sequences at the 5′ end of cellular pre-mRNAs (FIG. 1a), and splice sites influence the Pol II elongation rate and chromatin modification across the gene body (FIG. 1b). At the 3′ end of genes, Pol II pausing occurs after recognition of the polyadenylation site (PAS) by cleavage and polyadenylation factors[12] and owing to

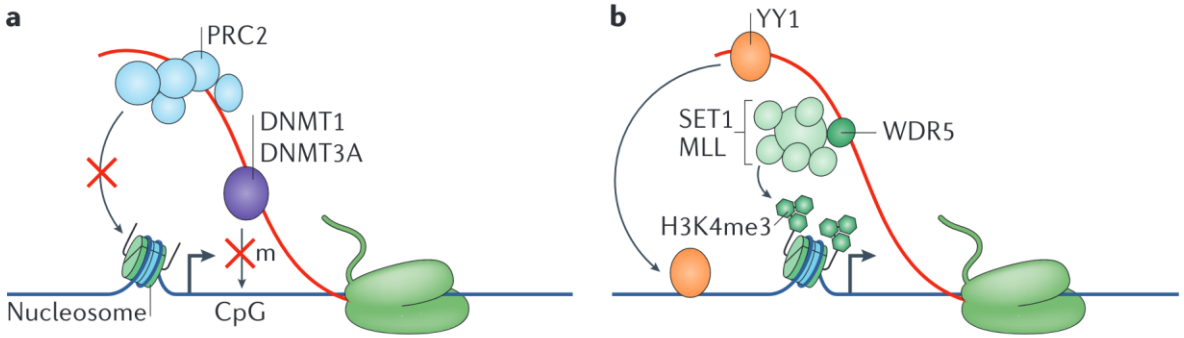


Figure 2 | **Nascent RNA modulates the association of regulatory factors with chromatin to maintain gene activity. a** | Nascent RNA can compete with chromatin for binding of repressive chromatin modifiers, such as Polycomb repressive complex 2 (PRC2), which methylates histone H3 at Lys27, and DNA (cytosine-5)-methyltransferase 1 (DNMT1) and DNMT3A, which primarily methylate the DNA at CpG dinucleotides. **b** | Interaction of the transcription factor yin and yang 1 (YY1) with nascent RNA facilitates its transfer to chromatin. Similarly, the interaction of WD repeat-containing 5 (WDR5), which is a component of the histone Lys methyltransferase complexes SET1 and myeloid/lymphoid or mixed-lineage leukaemia (MLL), with nascent RNA facilitates their transfer to chromatin and trimethylation of histone H3 at Lys4 (H3K4me3), thereby forming a positive-feedback loop that promotes gene expression.

Skalska, NRMCB, 2017

Nascent RNA is modified co-transcriptionaly

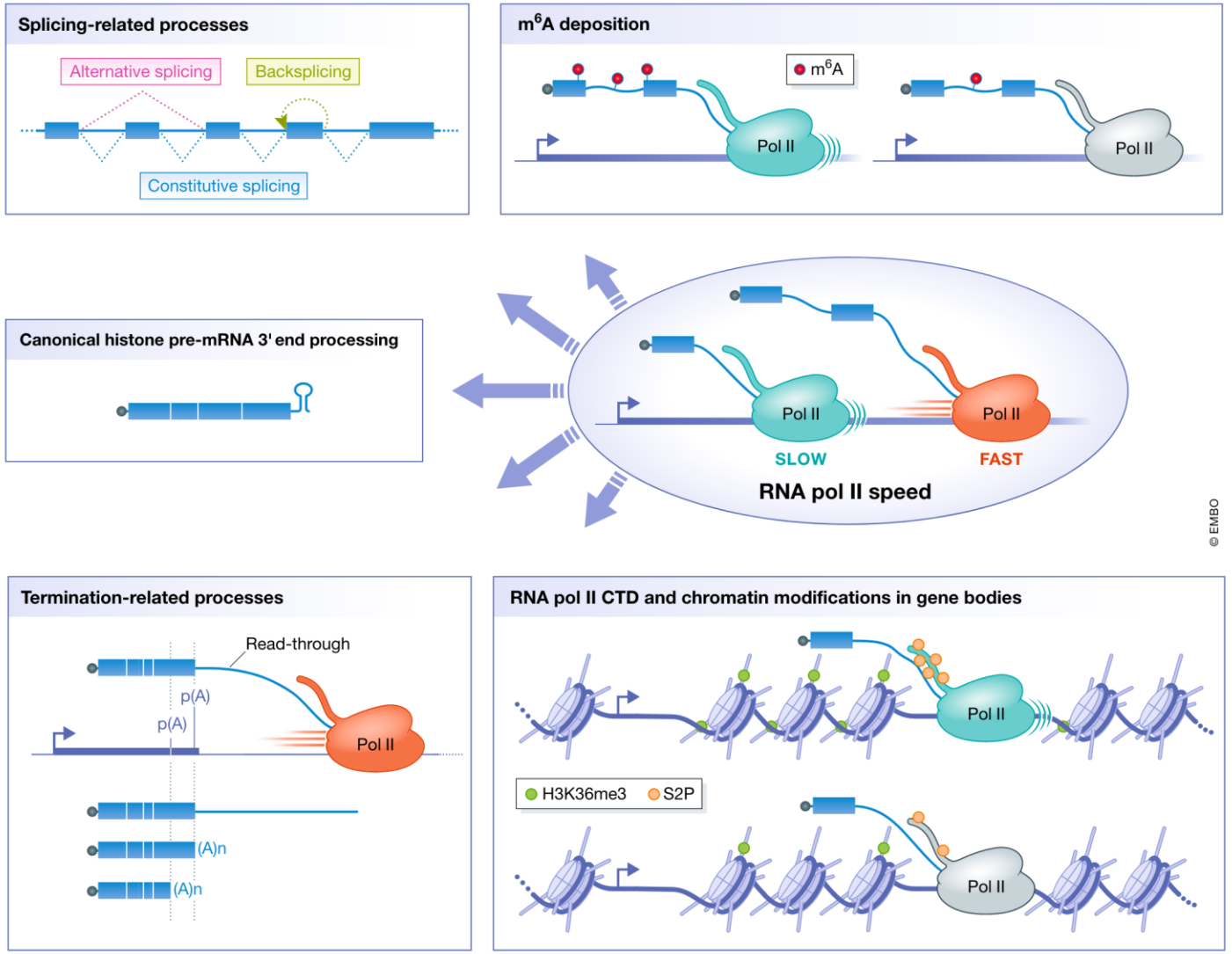


Figure 2. Co-transcriptional processes regulated by RNA Pol II speed.

Muniz et al, 2022

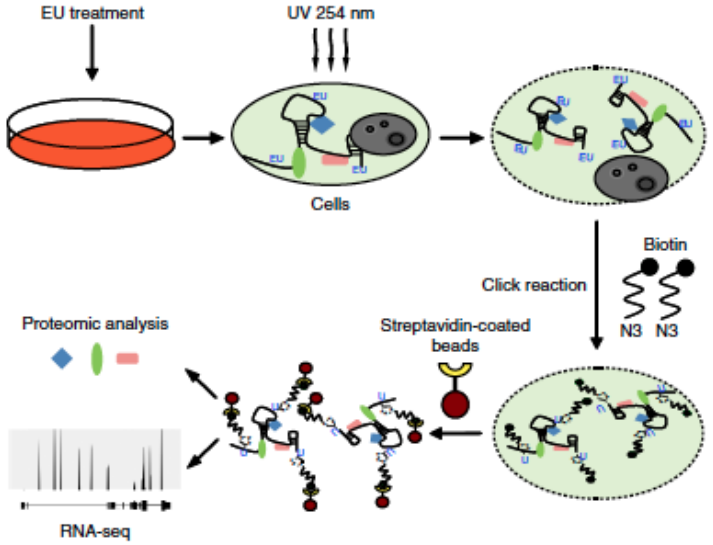
Emerging intriguing role of nuclear RNAs (including nascent pre-mRNAs) in shaping large-scale chromatin structure and regulating genome function

Capturing the interactome of newly transcribed RNA

Xichen Bao[1,2,24], Xiangpeng Guo[1,2,24], Menghui Yin[3,24], Muqddas Tariq[1,2,4], Yiwei Lai[1,2,4], Shahzina Kanwal[1,2], Jiajian Zhou[5], Na Li[1,2,6], Yuan Lv[1,2,4], Carlos Pulido-Quetglas[7], Xiwei Wang[1,2], Lu Ji[5], Muhammad J Khan[1,2,8], Xihua Zhu[1,2], Zhiwei Luo[1,2,4], Changwei Shao[9], Do-Hwan Lim[9], Xiao Liu[10], Nan Li[11], Wei Wang[12], Minghui He[13], Yu-Lin Liu[14], Carl Ward[1,2], Tong Wang[15], Gong Zhang[15], Dongye Wang[1,2,16], Jianhua Yang[17], Yiwen Chen[18], Chaolin Zhang[19], Ralf Jauch[16], Yun-Gui Yang[20], Yangming Wang[21], Baoming Qin[1], Minna-Liisa Anko[22], Andrew P Hutchins[23], Hao Sun[5], Huating Wang[5], Xiang-Dong Fu[9], Biliang Zhang[3] & Miguel A Esteban[1,2]

RECEIVED 19 JULY 2017; ACCEPTED 11 DECEMBER 2017; PUBLISHED ONLINE 12 FEBRUARY 2018; DOI:10.1038/NMETH.4595

NATURE METHODS | VOL.15 NO.3 | MARCH 2018 | 213



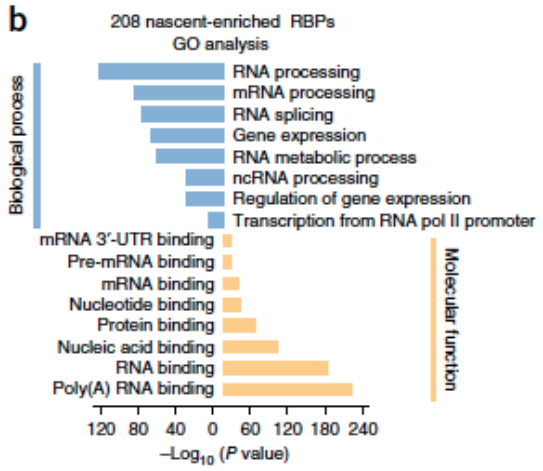
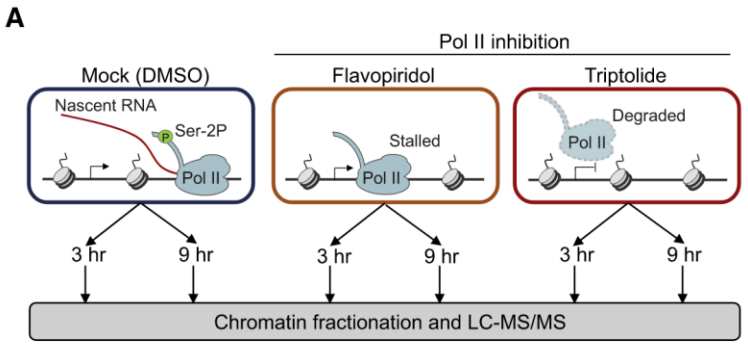


Figure 1 | Establishment of a new technique to capture the newly transcribed RNA interactome. Schematic representation of the RICK procedure.

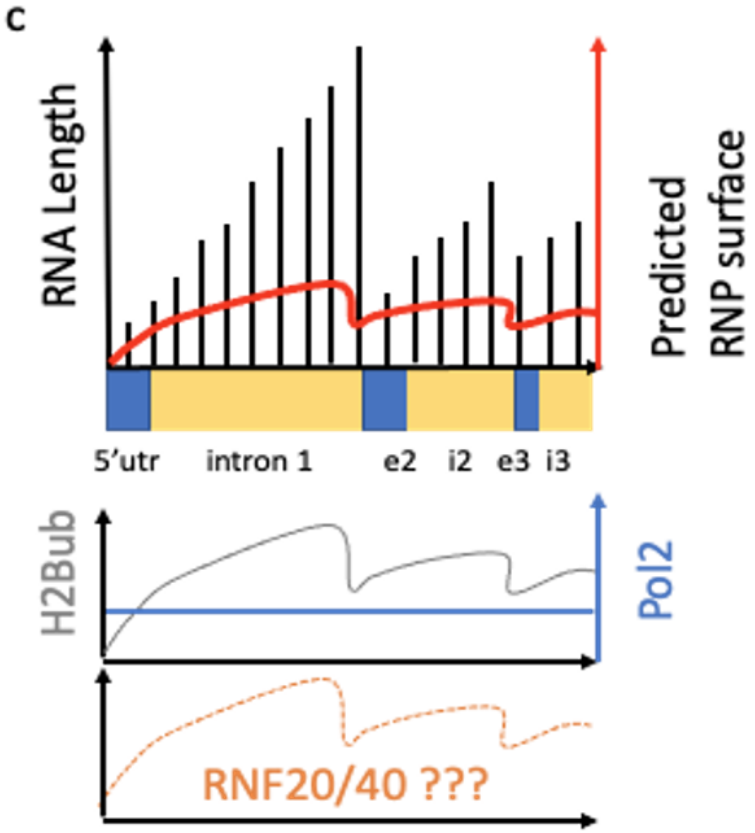
BONUS HIT: RNF20/40

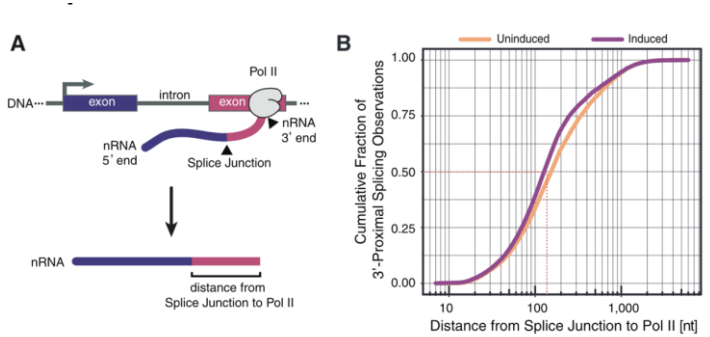


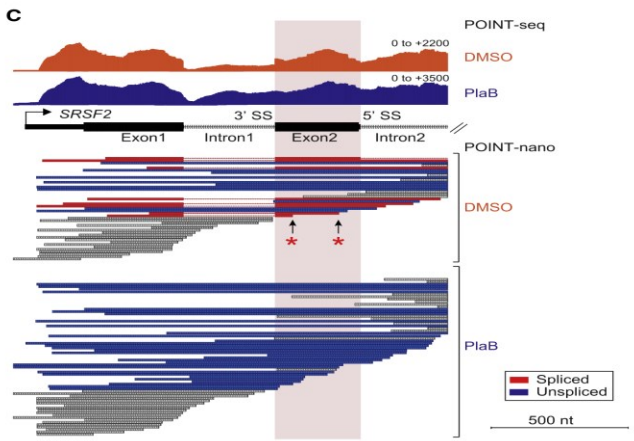
Human gene name	log2 FC vs mock (DMSO) control				
	Fp 3h	Fp 3h washou	Trp 3h	Fp 9h	Trp 9h
RNF20	-0.26	-0.22	-1.26	-0.17	-0.14
RNF40	-1.11	0.53		-0.61	0.06

Reanalysis of Skalska et al., 2021

Hypothesis: H2Bub writing depends on nRNA concentration/shape and is highly impacted by splicing



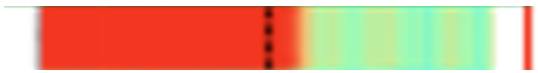




Analysis of LRS nRNA-seq

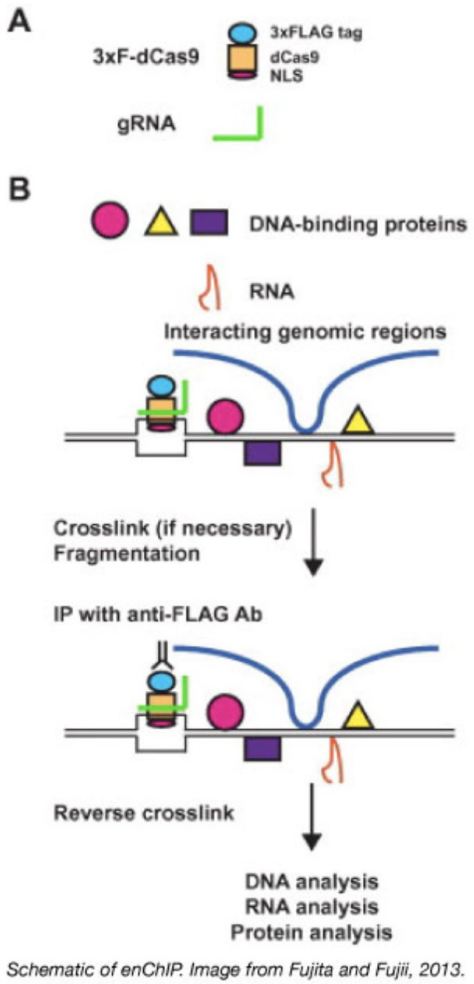
Exon start

H2Bub



Objective: *Determine co-transcriptional features distance RNAPII from Splice site and check differential splicing patterns and impact on h2bub/RNF20 levels*

Does splicing activity decrease the access of RNF20/40 in the following intron

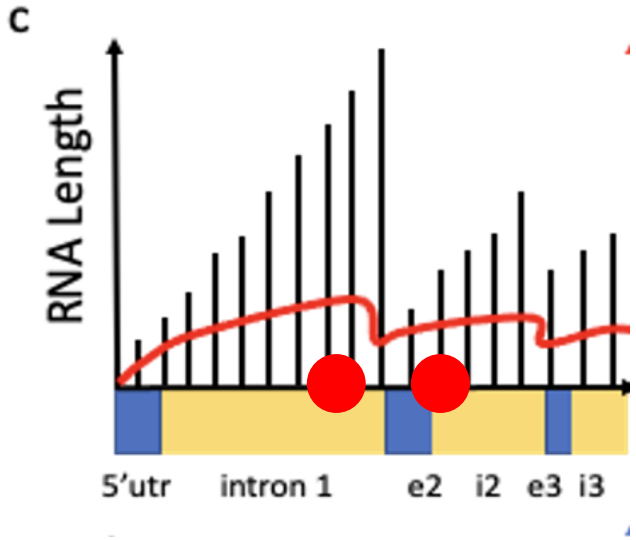


Schematic of enChIP. Image from Fujita and Fujii, 2013.

CRISPR-dCas9 to **pull-down** target loci **before or after splicing sites**:

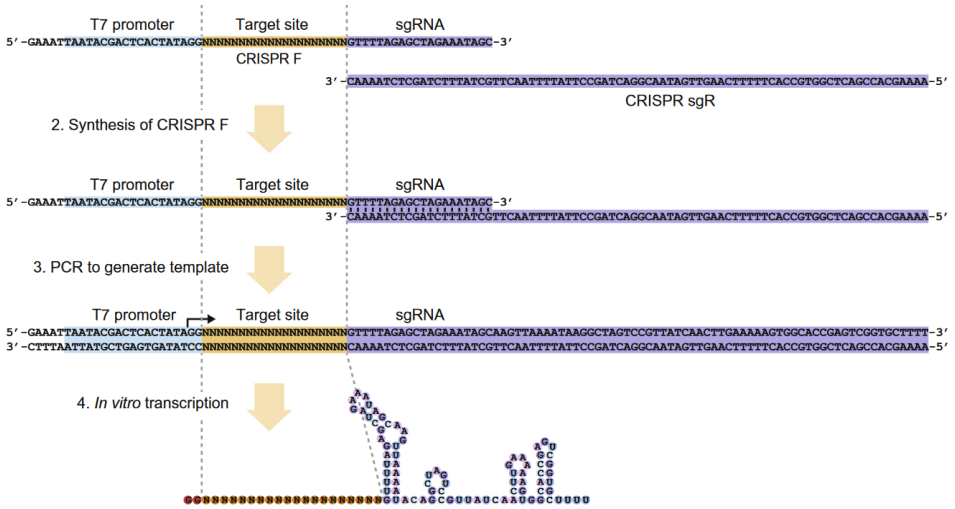
→ *analyze pulled-down vs input RNA and protein levels of H2Bub*

dCas9+gRNA

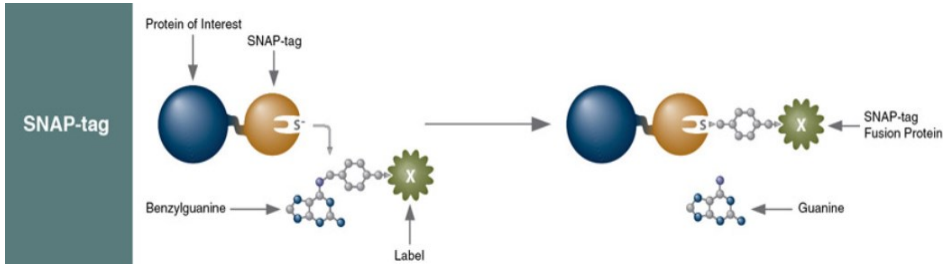


Objective : Define the structure of nRNA complexes tethered to Pol2 and the impact of splicing on H2Bub writing

Crispr pull down system (In vitro) to avoid transcription arrest in vivo

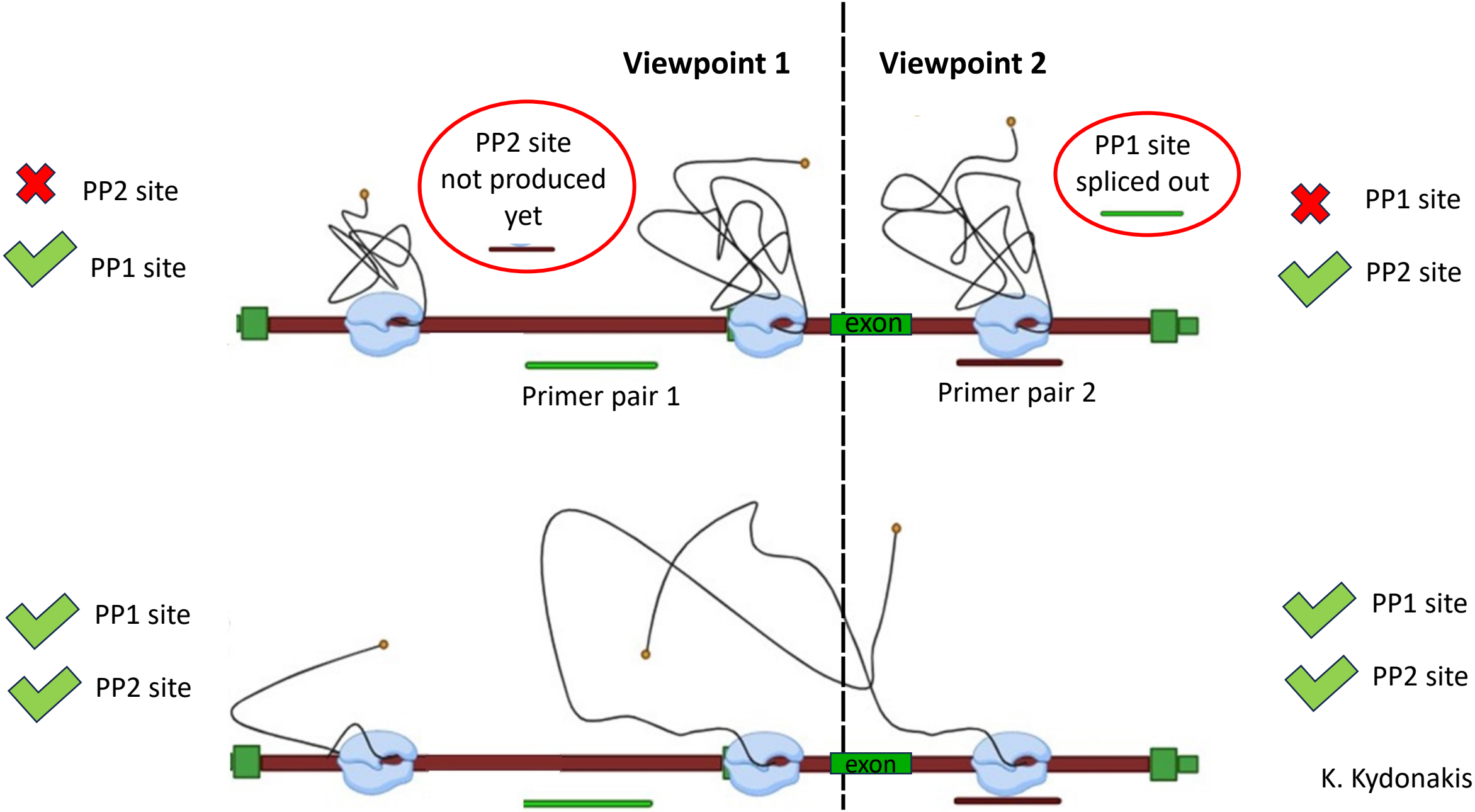


CRISPR/Cas9 mediated genome engineering in Drosophila 10.1016/j.ymeth.2014.02.019



- Suicide enzyme (Single action)
- Irreversible covalent binding
- SNAP-tagged dCas9
- Elution by Proteinase K

CRISPR PULL DOWN

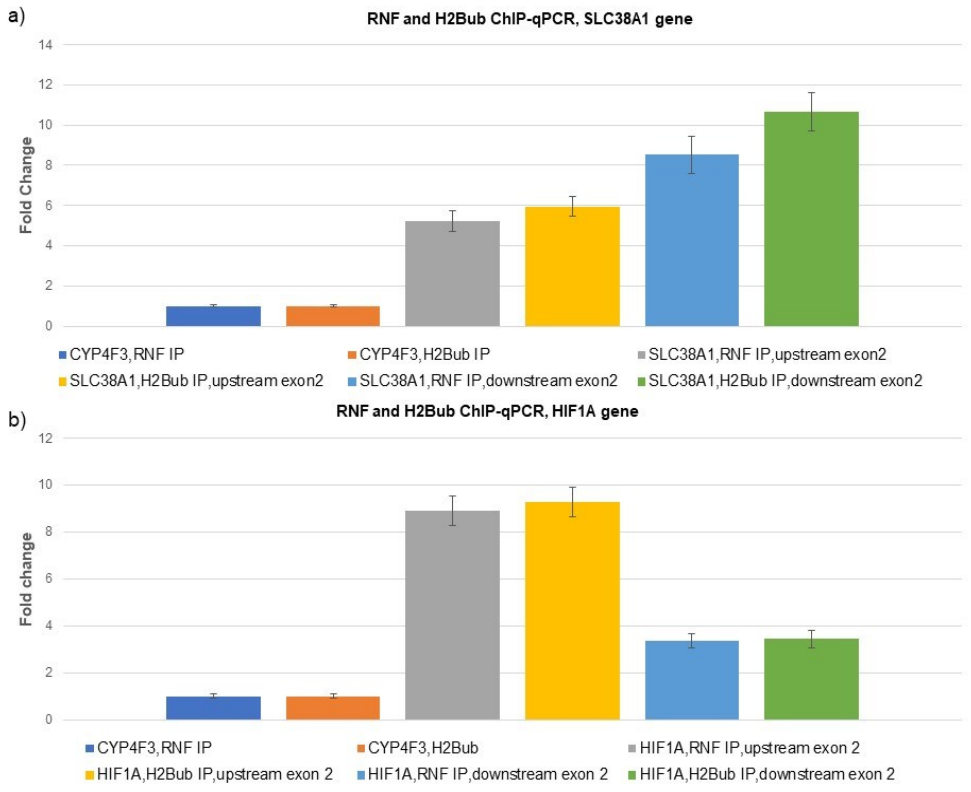


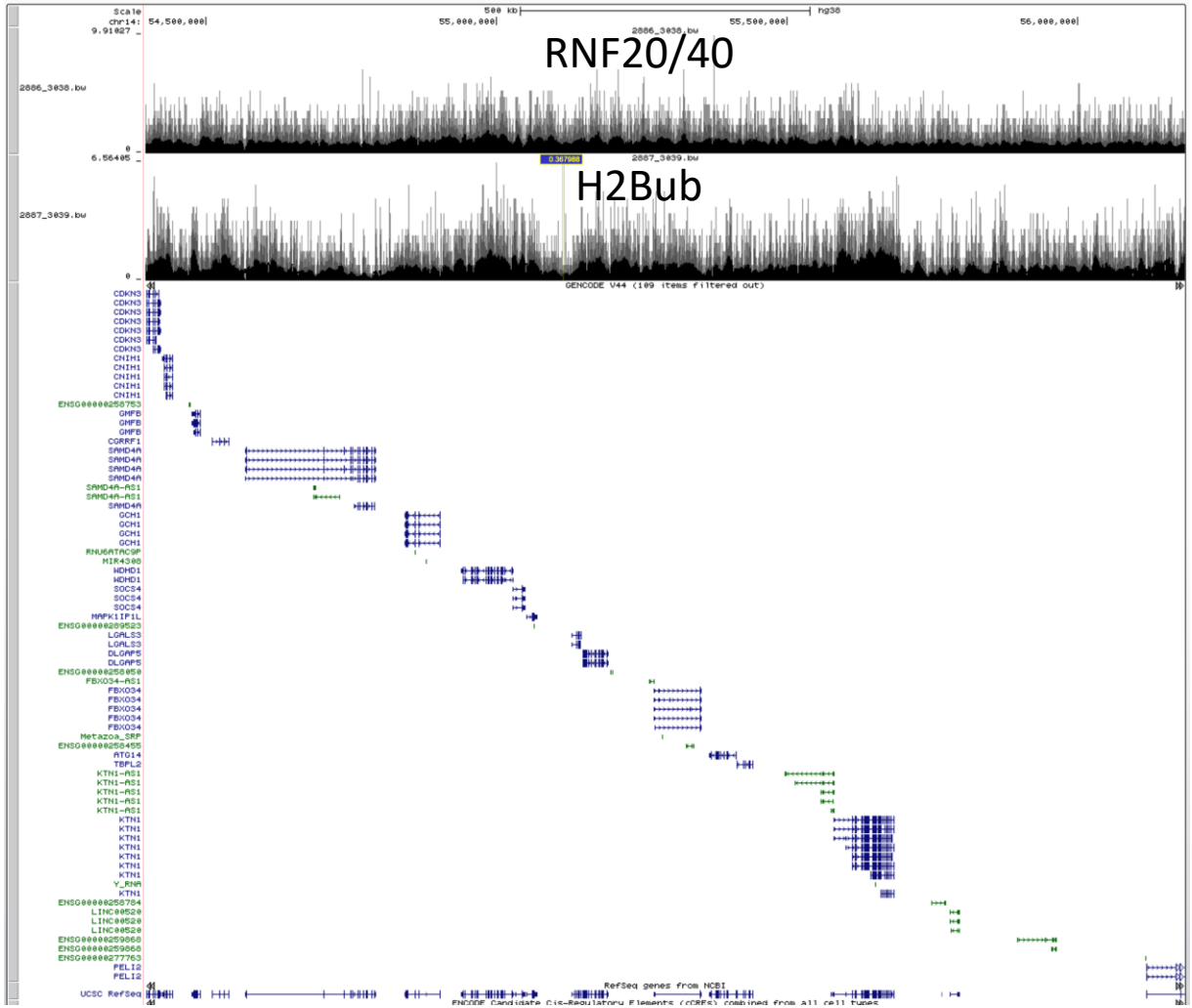
K. Kydonakis

Establish patterns of RNF20/40 on chromatin

Map RNF20/40 vs H2Bub in genome or or nRNA

anti-RNF20
+
antiRNF40





V. Theodosiou

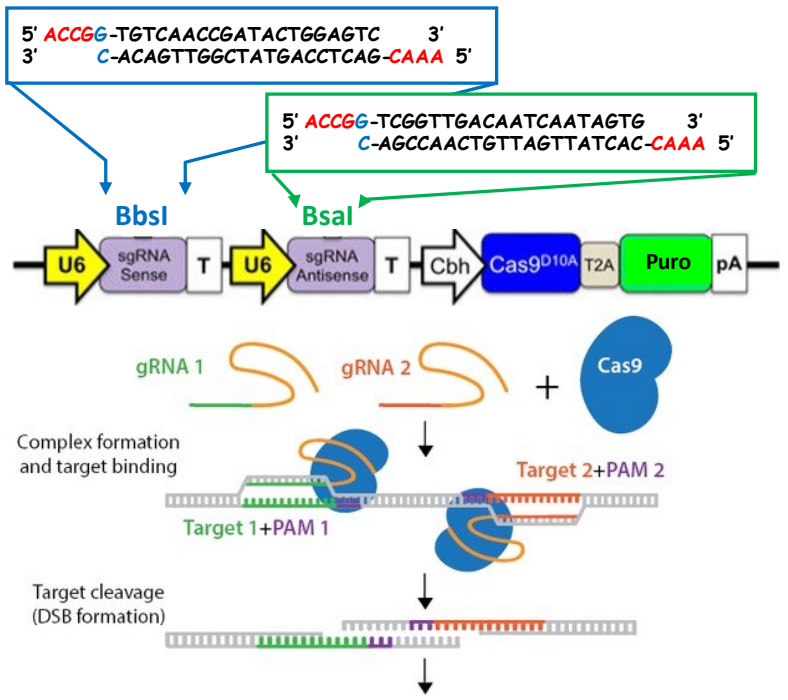
Objective : TWO birds with one stone!!!

Produce inducible degradation of RNF20/40 HA **tagged proteins for efficient ChIP/RIP-seq and functional studies**

Study RNF20/40 loss-of-functions in cancer cell lines

(Vaios Theodosiou et al)

a



DSB repair by NHEJ disrupts RNF20 frame --> **KO**

b

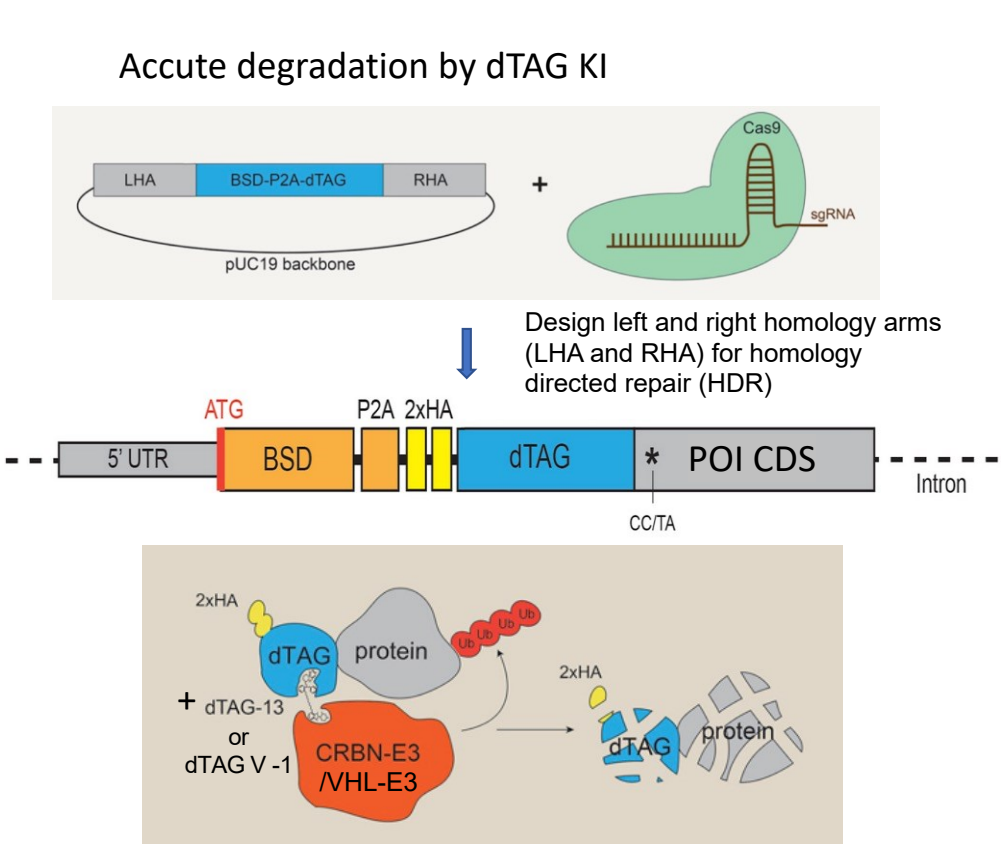


Figure 4: Strategies to generate modified cell lines. **a-** KO by CRISPR-Cas9(D10A) and **b-** acute depletion of POI by dTAG KI. POI: Protein of Interest, DSB: Double Strand Break.

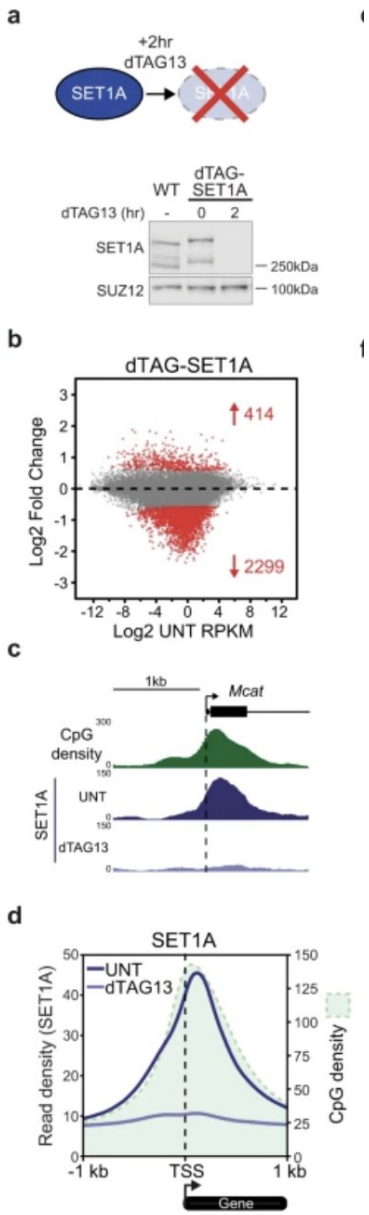
Objective: generate and analyse RNA-seq, ChIP-seq, GRID-seq, Promoter-HiC and ATAC-seq in the dTAG cell lines

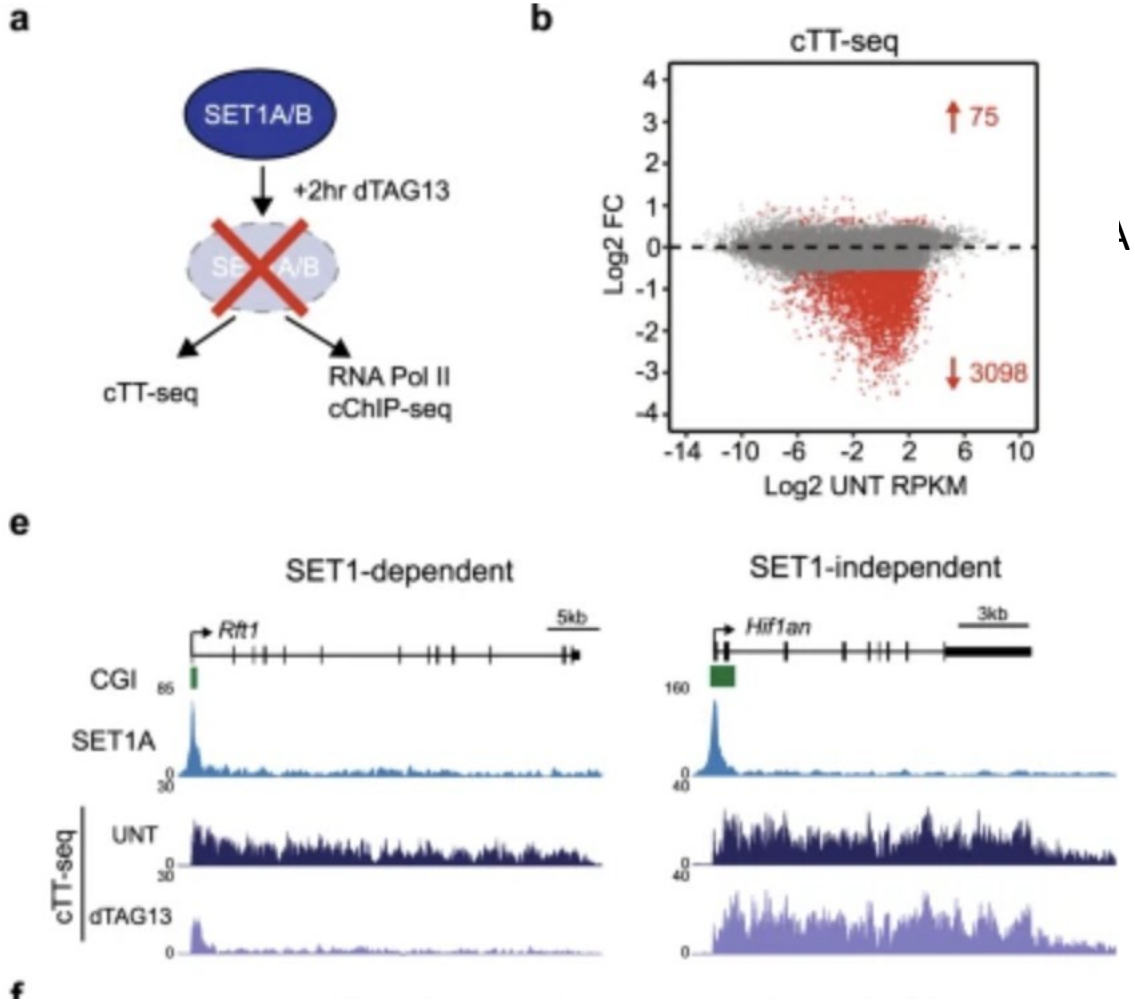
dTAG very efficient to study chromatin processes

A CpG island-encoded mechanism protects genes from premature transcription termination

Amy L. Hughes, Aleksander T. Szczurek, Jessica R. Kelley, Anna Lastuvkova, Anne H. Turberfield, Emilia Dimitrova, Neil P. Blackledge & Robert J. Klose

Nature Communications **14**, Article number: 726 (2023) | Cite this article

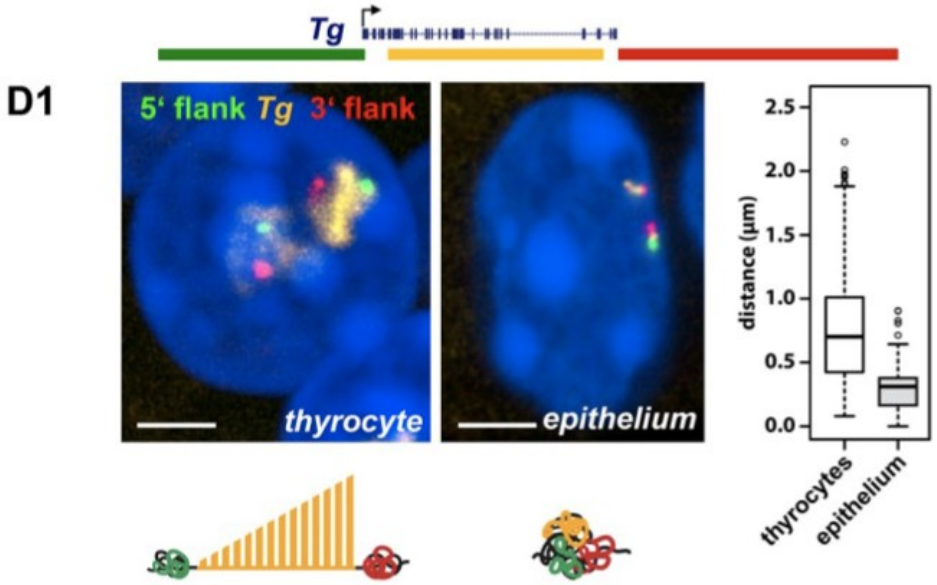


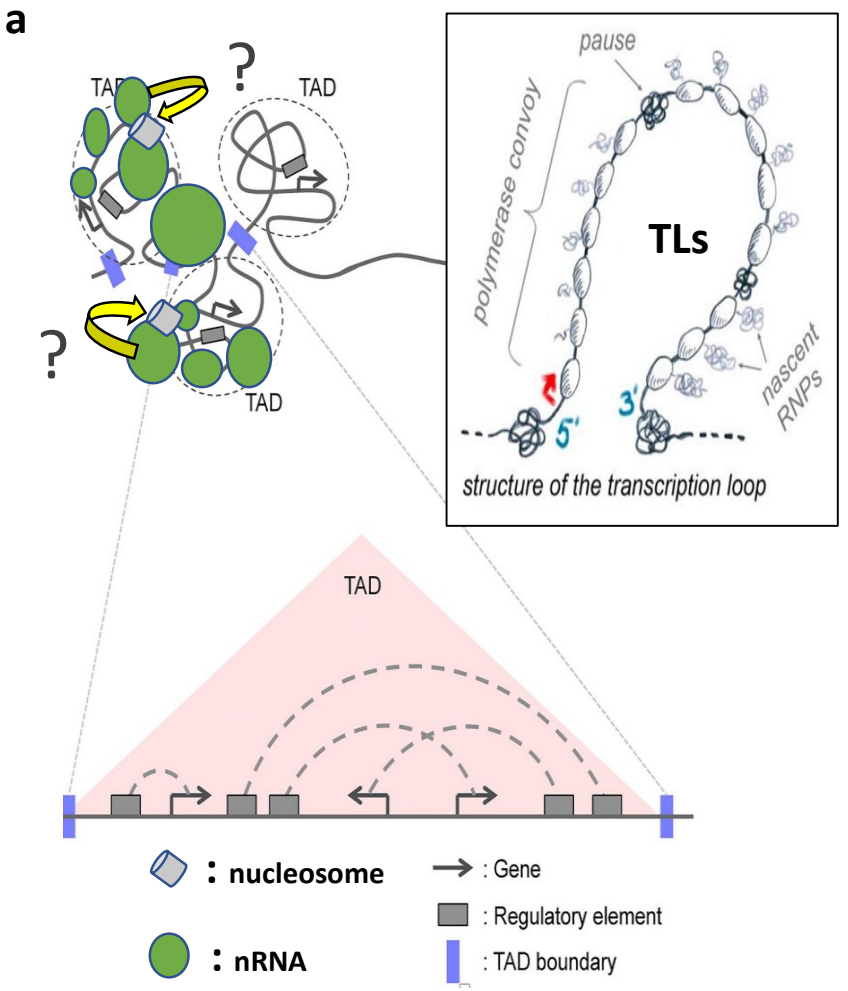


Objective: Understanding how **nascent RNA** participates in global genome/chromosome conformation dynamics

Objective: study nascent RNA shape/interactions and determine effect of splicing on nascent RNA volumes

Study RNA-DNA 3D contacts
→ Infos on RNPs structure at given genomic loci
→ nRNA role in chromatin contacts
→ **Perturbations in stress cancer???**



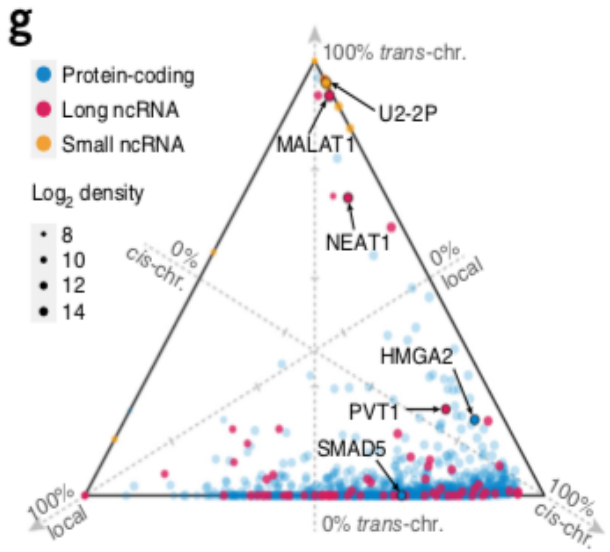


Modified from Leidedscher, NCB, 2022

Role of nascent RNA in stabilizing chromatin loops and transcriptional hubs?

(Kostis Kydonakis et al.)

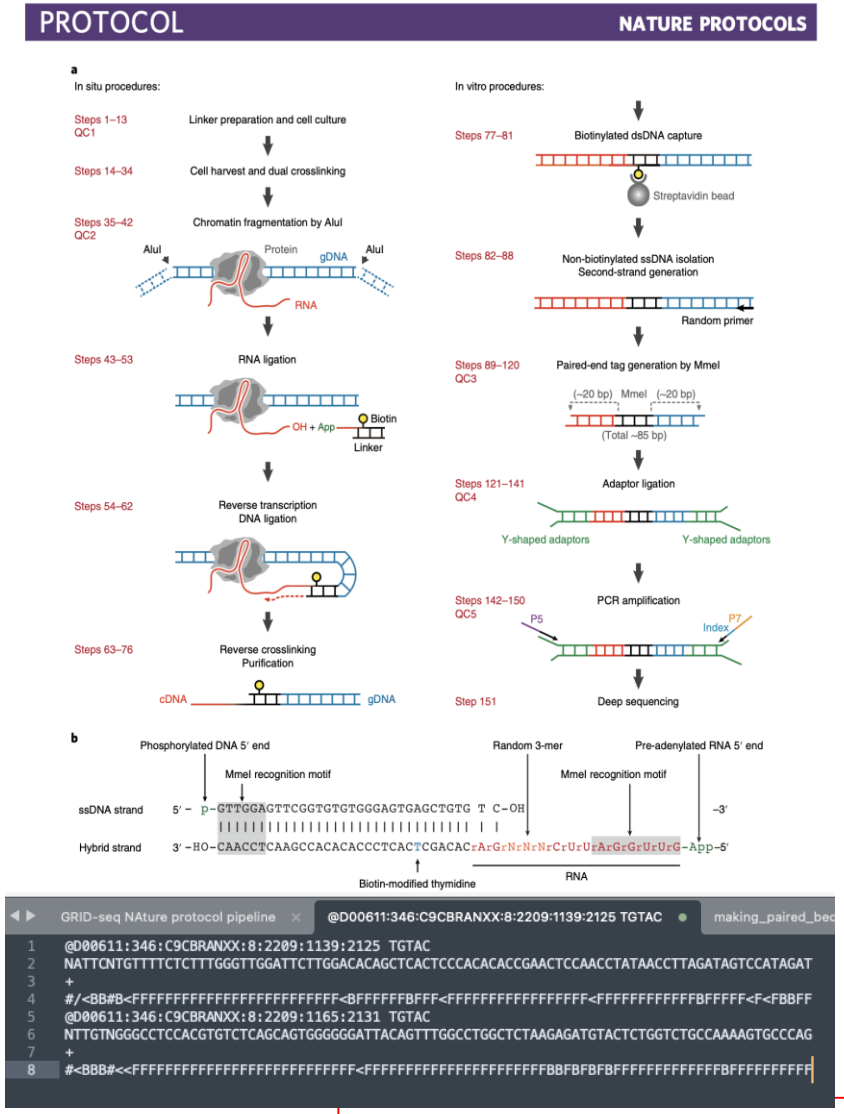
GRID-seq (NB, 2017)



We USE IT FOR: Study RNA-DNA 3d contacts

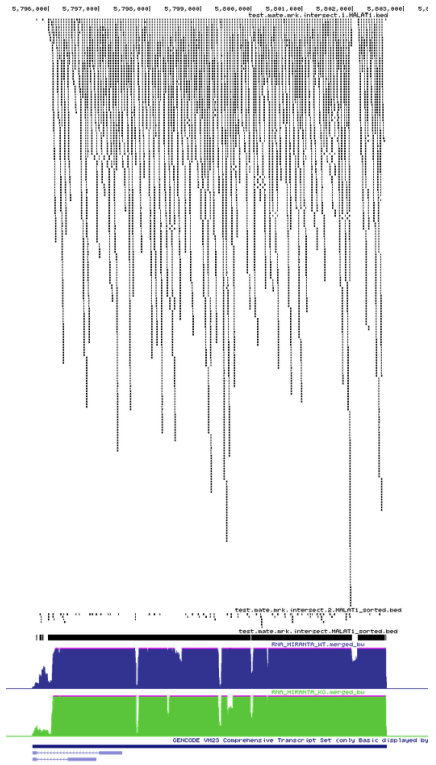
→ Infos on RNA structure at given genomic loci

--> nRNA role in chromatin contacts



GRID RNA

DNA



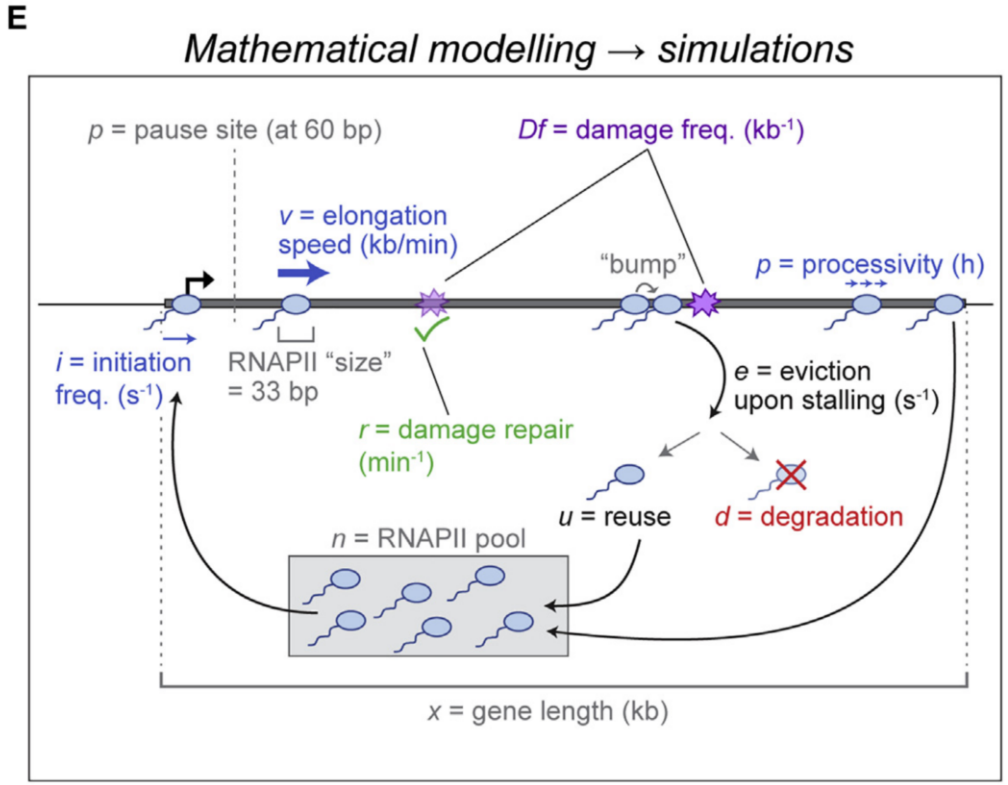
analsyse GRID-seq data to understand nRNA interactome

Check the DNA vs RNA around intron-exon juntion

Objective :

Understanding transcription regulatory networks by Machine learning mathematical modeling and simulation

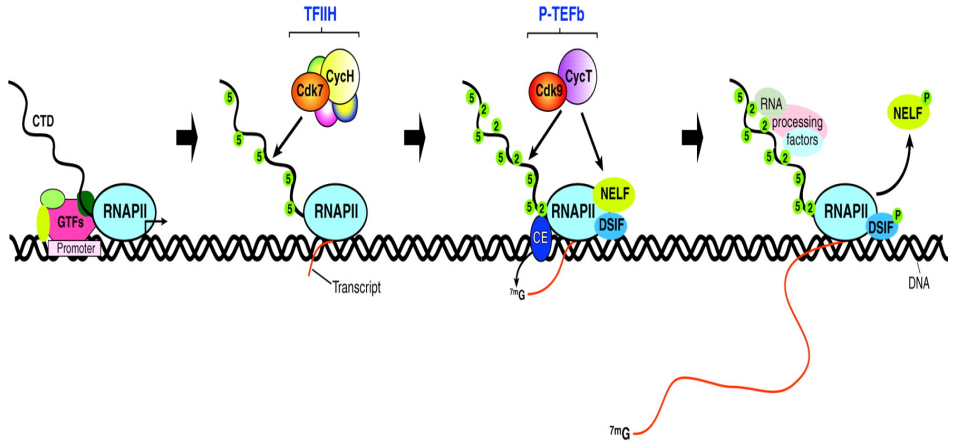
Simulating in-silico regulatory steps of transcription initiation, pausing, elongation, termination



Svejstrup lab

RNAPII C-Terminal Domain (CTD) status

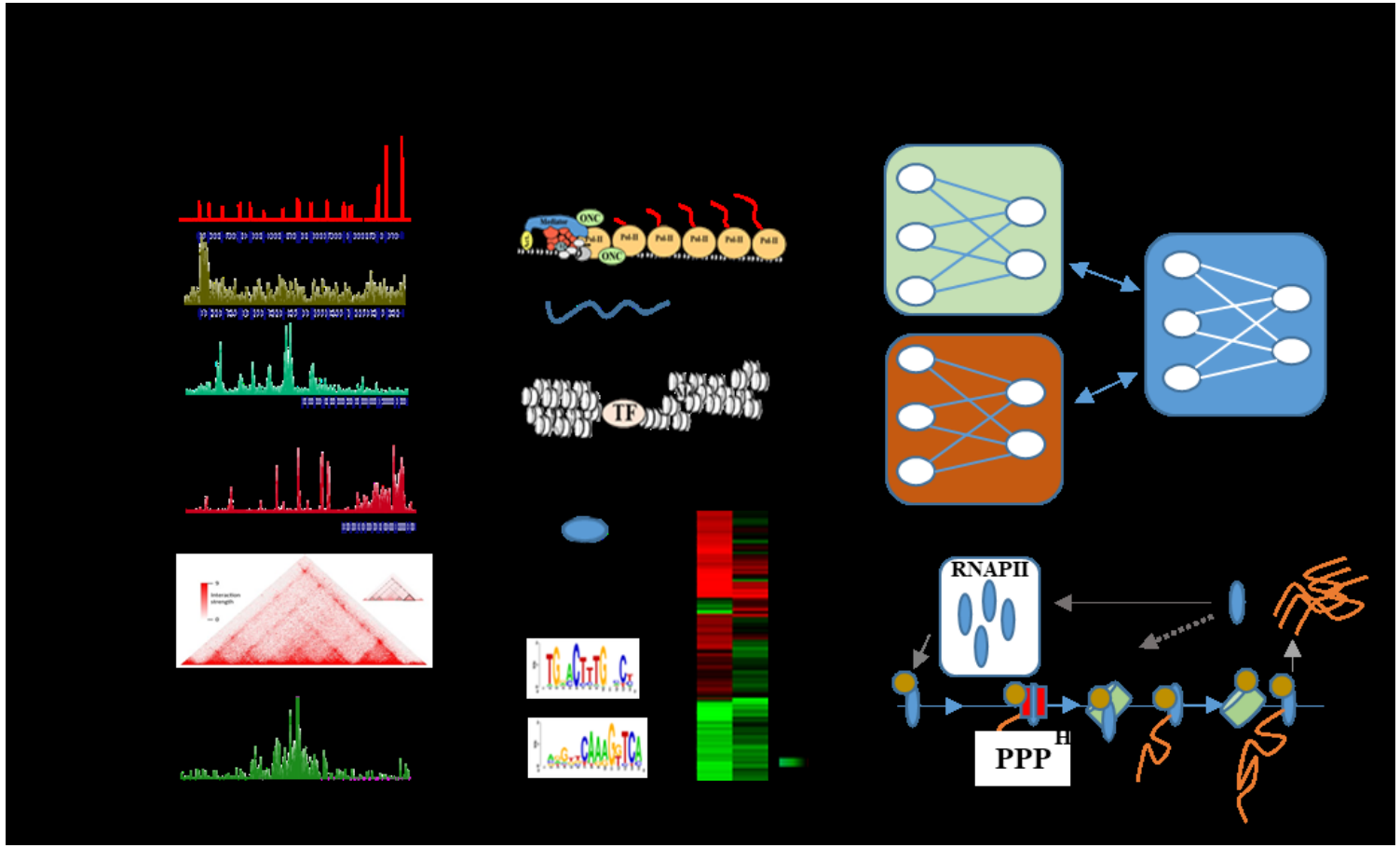
PIC	Initiation	PPP= Promoter-Proximal Pausing	Elongation
Hypo ser5P		Signal-dependent transition into elongation	ser2P



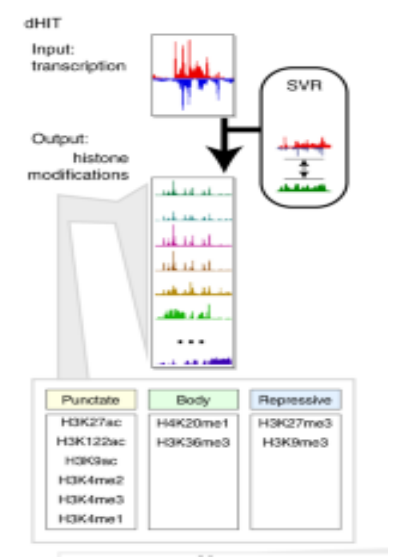
TSS PPP

Mathematical modeling:

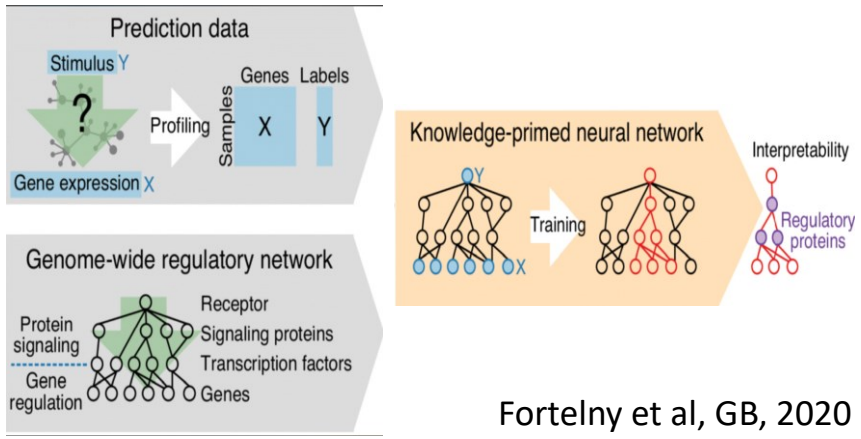
identify the principles of how the different parameters of the distinct steps of the transcription process can predict the occurrence of the others and determine actionable targets for cancer therapy



ML to infer mechanisms of regulation



Wang et al, NG, 2022



Fortelny et al, GB, 2020

Collaboration with Talianidis lab (IMBB)

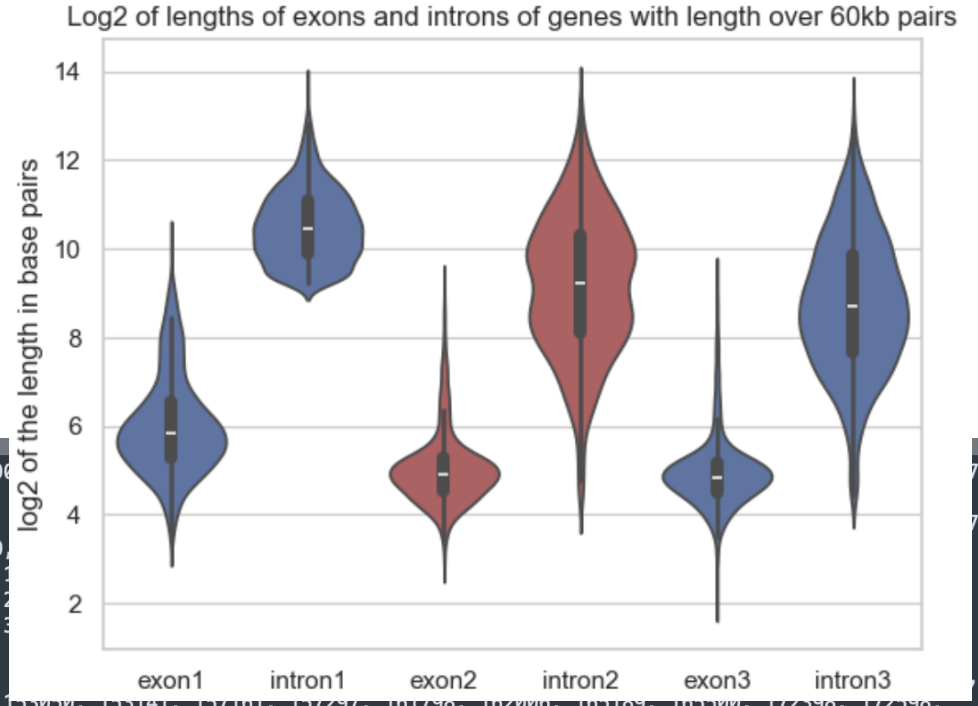
Preparing annotation of gene structures for Simulator

Python script to filter genes based on features

```
if (gene_length != 'optional' and intron_length_required != 'optional'):
    data = pd.read_csv(file_name, sep = '\t')
    dataframes_of_different_genes = [y for x, y in data.groupby('gene_name')]

    with open(output_file_name, 'w') as out:

        for gene_id in range(len(dataframes_of_different_genes)):
            gene_exons_length_dict = defaultdict(int)
            number_of_exons = len(dataframes_of_different_genes[gene_id])
            Name_of_the_gene = dataframes_of_different_genes[gene_id]['gene_name'].values[0]
            length_of_the_gene = dataframes_of_different_genes[gene_id]['gene_length'].values[0]
            if ',' in length_of_the_gene:
                length_of_the_gene = length_of_the_gene[:length_of_the_gene.index(',')]

            for i in dataframes_of_different_genes[gene_id].index:
                gene_exons_length_dict[int((dataframes_of_different_genes[gene_id]['exon_id'][i])[dataframes_of_different_genes[gene_id]['exon_id'][i].index('e')+1:])] =  (dataf
            gene_exons_length_dict_sorted_by_exons =  dict(sorted(gene_exons_length_dict.items()))
            print(gene_exons_length_dict_sorted_by_exons)


            if (int(length_of_the_gene) >= gene_length) :
                intron_length = []
                for i,j in list(zip(list(gene_exons_length_dict_sorted_by_exons.values())[:-1],list(gene_exons_length_dict_sorted_by_exons.values())[1:])):
                    intron_length.append(j[0] - (i[1]))
                intron_length.append('end')
                exon_length_intron_length = list(zip(list(gene_exons_length_dict_sorted_by_exons.values()), intron_length))
                if intron_length[0] > intron_length_required:
                    to_print = []
                    for i in list(gene_exons_length_dict_sorted_by_exons.values()):
                        if i[0]-list(gene_exons_length_dict_sorted_by_exons.values())[0][0] >= 0 :
                            to_print.append((i[0]-list(gene_exons_length_dict_sorted_by_exons.values())[0][0], i[1]-list(gene_exons_length_dict_sorted_by_exons.values())[0][0]))
```

AIM: to be able select genes to analyse based on:

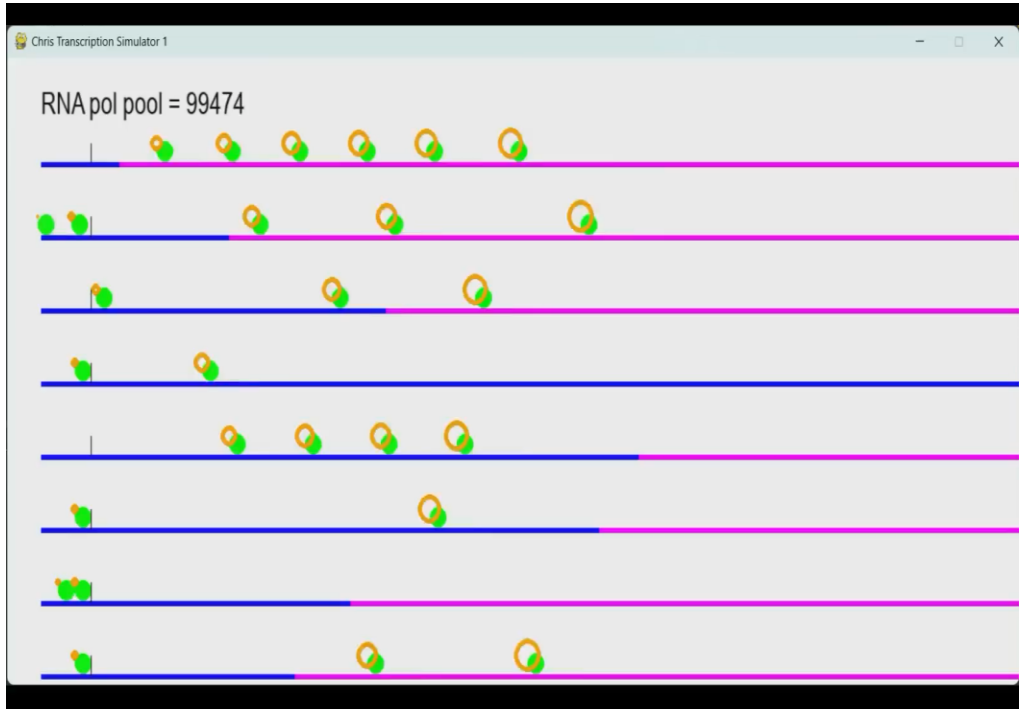
gene legth
intron length
Differential expression patterns,
Alternative splicing patterns




```
ABAT,71607,0, 125, 61112, 61223, 71414, 71512, 73521, 73551, 75835, 75953, 83170, 83220, 89482, 89563, 90
101757, 101904, 104892, 105004, 106722, 109989
ABCA13,476035,0, 93, 21499, 21593, 26777, 26901, 47894, 48046, 49821, 49850, 55802, 55966, 58365, 58496,
76779, 76985, 77752, 77892, 97520, 97635, 100327, 105106, 106634, 108461, 109883, 109993, 116500, 116619,
142772, 142950, 145697, 145837, 163962, 164076, 166905, 166996, 179183, 179360, 180721, 181028, 196334,
217761, 220461, 220680, 232223, 232420, 239060, 239218, 240893, 241124, 256306, 256412, 283577, 283827,
334875, 334930, 336411, 336589, 339623, 339739, 345264, 345421, 348580, 348834, 352787, 352980, 356775,
471933, 473156, 473294, 473956, 476035
ABCC1,193497,0, 223, 58239, 58416, 60199, 60325, 64914, 65052, 66919, 67045, 83532, 83594, 86895, 87027,
118726, 122065, 122153, 126749, 126825, 129775, 129902, 133789, 133966, 137247, 137415, 140828, 141012,
175212, 175339, 176233, 176335, 182212, 182359, 184773, 184932, 186901, 187068, 188787, 188982, 191596, 193497
```

Nascent RNA role in transcription regulation and chromatin conformation

Simulation of:
RNAPII dynamics
co-transcriptional processes
nascent RNA volume change at intron-exon-junctions



Virtual Genome: genes with 1st intron > 10kb

C. Botos, C. Zadmirah and M. Stagaki

PROJECT: Design Simulators (Python, C) of TRANSCRIPTION /CO-SPLICING PROCESSES

Collaboration with Dr Katsaounis (applied mathematics)

Acknowledgements

Vaios Theodosiou (PhD Student)
Electra Tsaglioti (RA Bioinfo)
Marianna Stagaki (*MSc Bioinfo thesis*), Angeliki Loukopoulou (Msc Rotator), Maria-Electra Kontonikou, Chris Botos (BSc Biology thesis)
Kostis Kydonakis, Myrto Mittleton (MSc thesis)
Nikos Vouzounerakis, Stergios Manakas, John Petrossian
Nektaria Kokolaki, Nektarios Belmezos, (BSc Biology thesis)
Katerina Papadaki (BSc Math project)
Chadmirah Zaratiana (ERASMUS, University of Paris)
Dusanka Lumovic, Sofia Kaforou (Technicians)

Hiring!!!

bench
bioinfo

Collaborators:

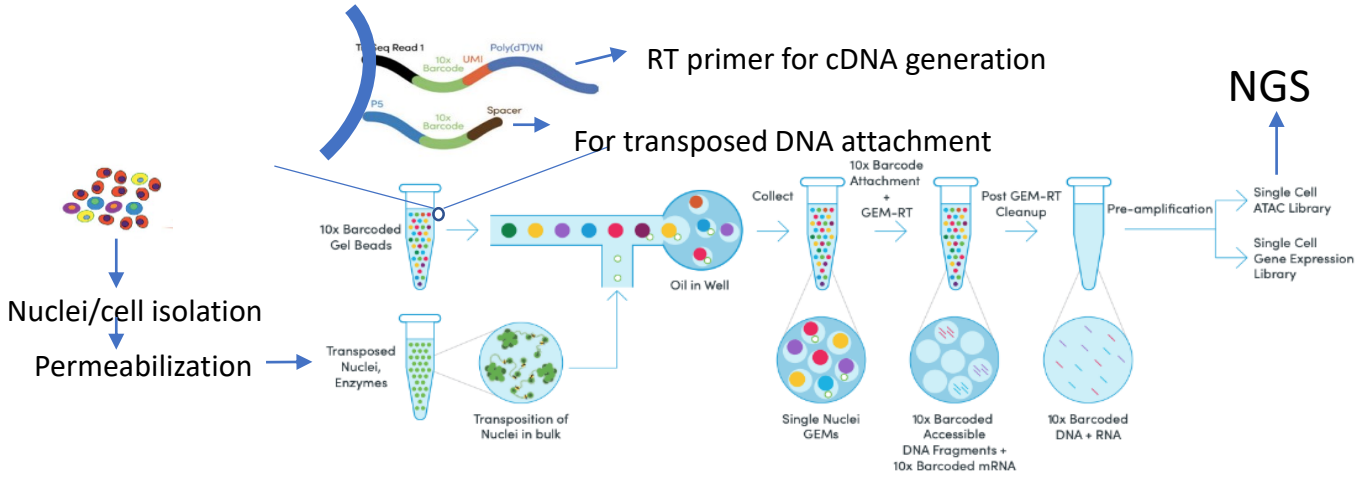
ITE
Vasso Theodorou and Christos Delidakis
Talianidis lab
Ntini lab
Garinis lab
Pavlopoulos lab
Verginis lab
Chamilos lab
Charalampopoulos lab
Kalantidis lab
Bertsias lab
Katsaounis lab (IACM)
Filippidi lab (IESL)
Genomics and Cell culture facilities

Fleming
Fousteri lab

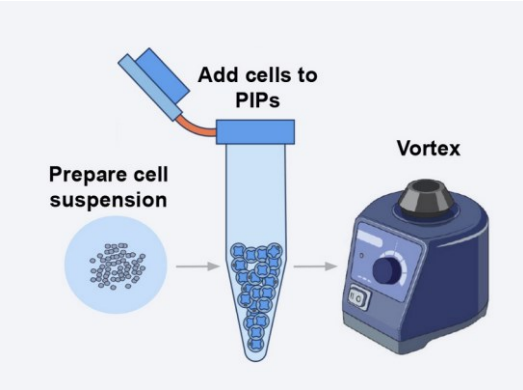
FORTH INSTITUTE OF MOLECULAR BIOLOGY & BIOTECHNOLOGY
H.F.R.I. Hellenic Foundation for Research & Innovation
FORTH Synergy
Ελλάδα 2.0 ΕΘΝΙΚΟ ΣΧΕΔΙΟ ΑΝΑΚΑΜΨΗΣ ΚΑΙ ΑΝΘΕΚΤΙΚΟΤΗΤΑΣ

e-mail: lavigne@imbb.forth.gr, @matlavmac

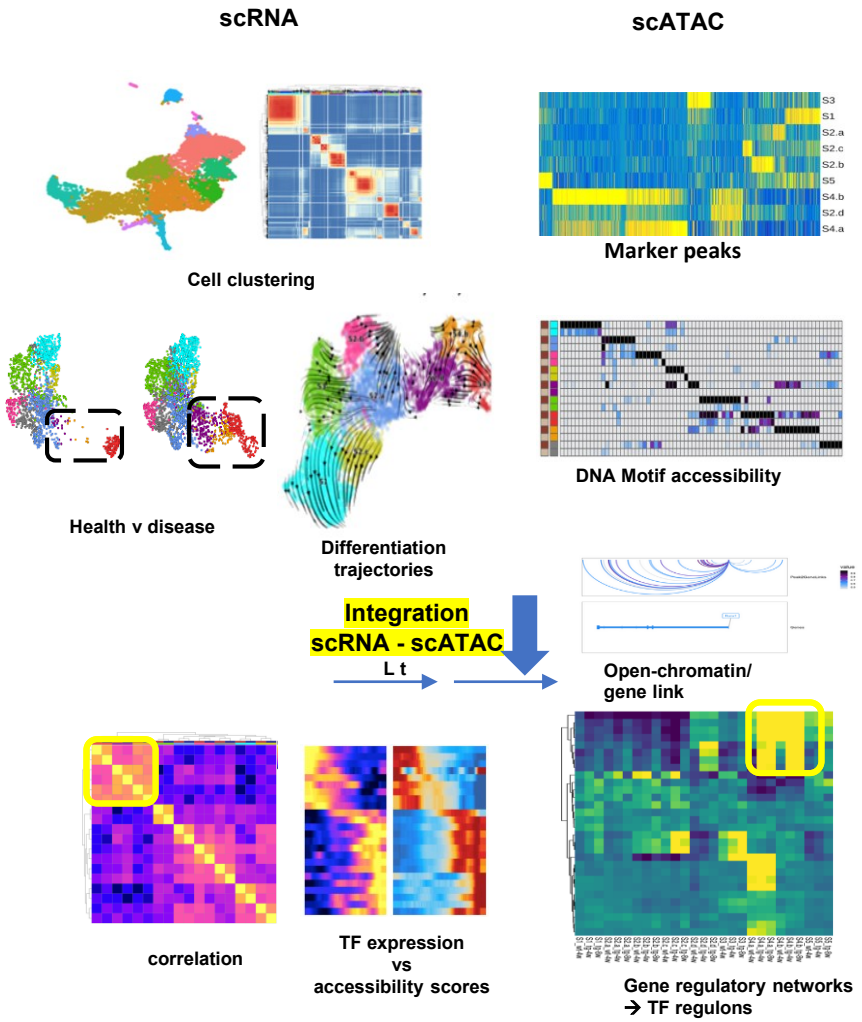
Integrative single-cell (sc) analysis of chromatin and transcriptome dynamics to investigate gene regulatory bases of Disease



10x sc multiome ATAC+ GEX workflow



PIP-seq (Fluent Biosciences with ENZYQUEST)



Methodolgy applied in i) Armaka et al, 2022, ii) Peraki and Botskaris, ms in preparation (in collaboration with Talianidis lab)

Objective 3: Investigating the MOLecular SPEcificity of chromatin CONDensates
MOLSPECOND (FORTH Synergy grant)

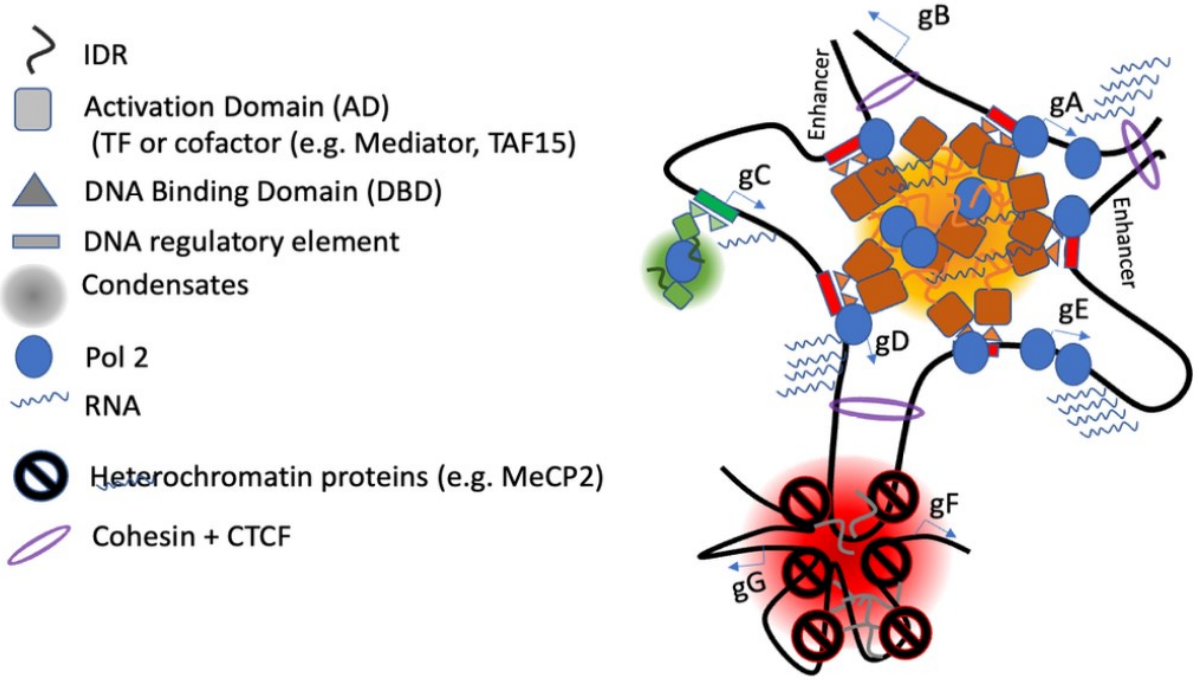
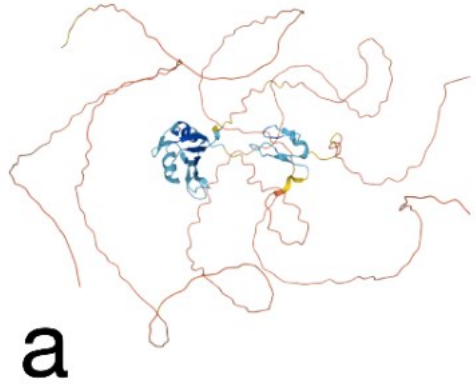


Figure 1: Multivalent interactions lead to the formation of condensates localized to specific genomic regions. A condensate of two enhancers (eRNA is a scaffold to phase separation) and associated promoter concentrate pol 2 (blue) and coactivators (squares) via the interaction of their IDRs and DBDs to increase transcription. Cohesin extrudes DNA into loops until it encounters occupied CTCF DNA-binding sites and participate to the topology of chromatin condensates. Heterochromatin condensates prevent transcriptional activity and segregate inactive regions of the genome out of active sites.



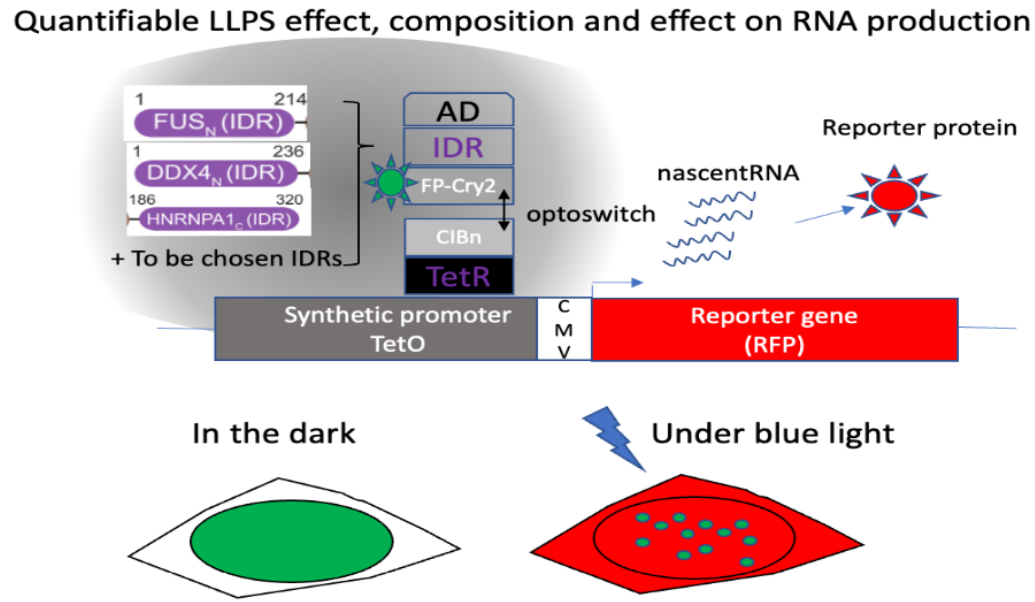


Figure 3: Workflow of in vivo reporter assay to measure the impact of IDR sequence variability on condensate composition upon blue light (Cry2-CIBn) optogenetic switch. Plasmids obtained from Schneider et al, 2021 necessary for expression of all components will be modified to suit our goal to test colocalization and effects on transcription output.

Could Pol II and **active transcriptional processes** be key drivers mediating finescale functional chromatin structures?

nature genetics

Article https://doi.org/10.1038/s41588-023-01364-4

Enhancer–promoter contact formation requires RNAPII and antagonizes loop extrusion

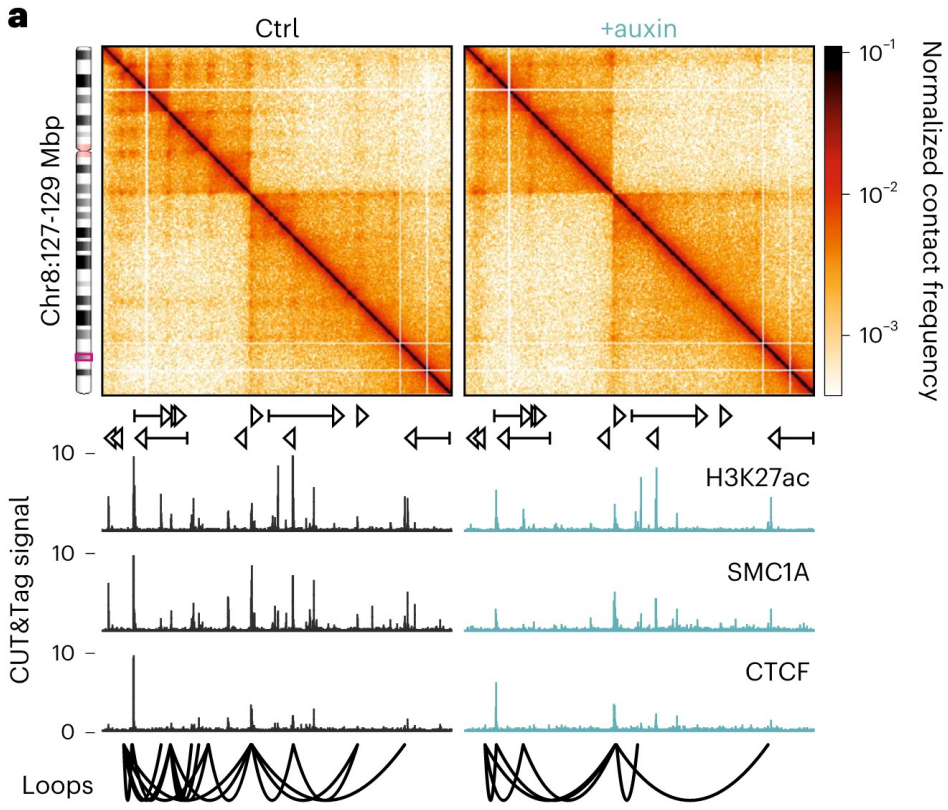
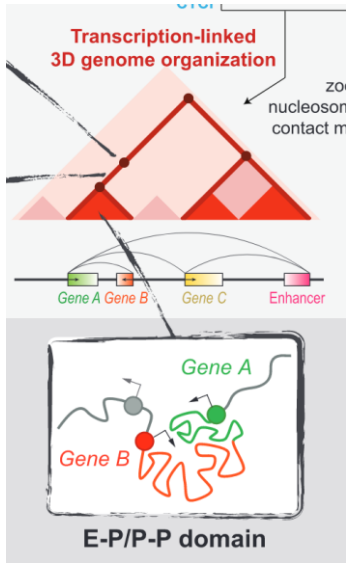
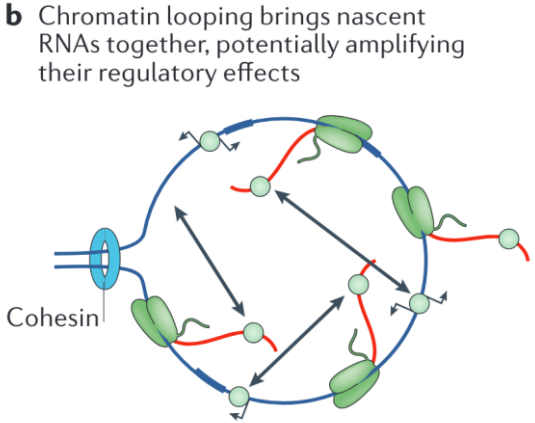
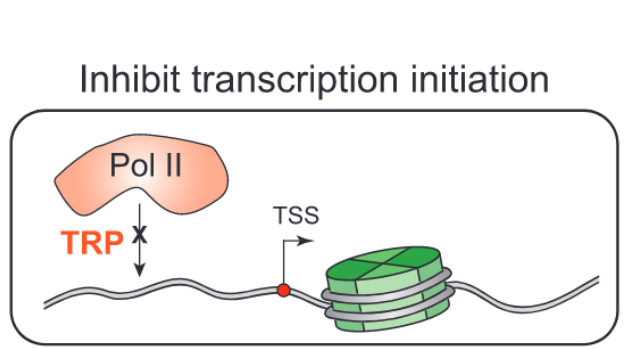


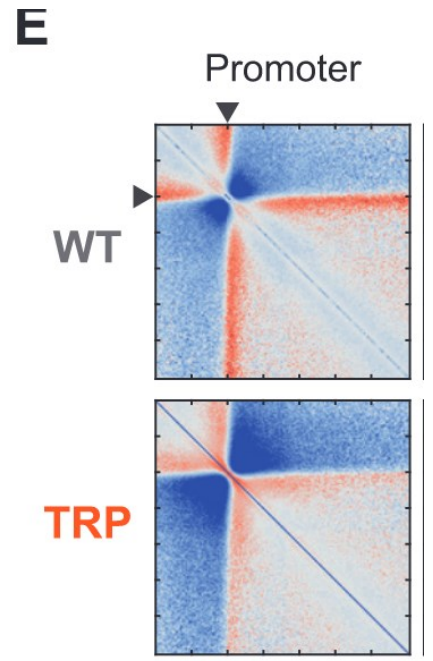
Fig. 3 | RNAPII depletion selectively affects enhancer–promoter and enhancer–enhancer loops.





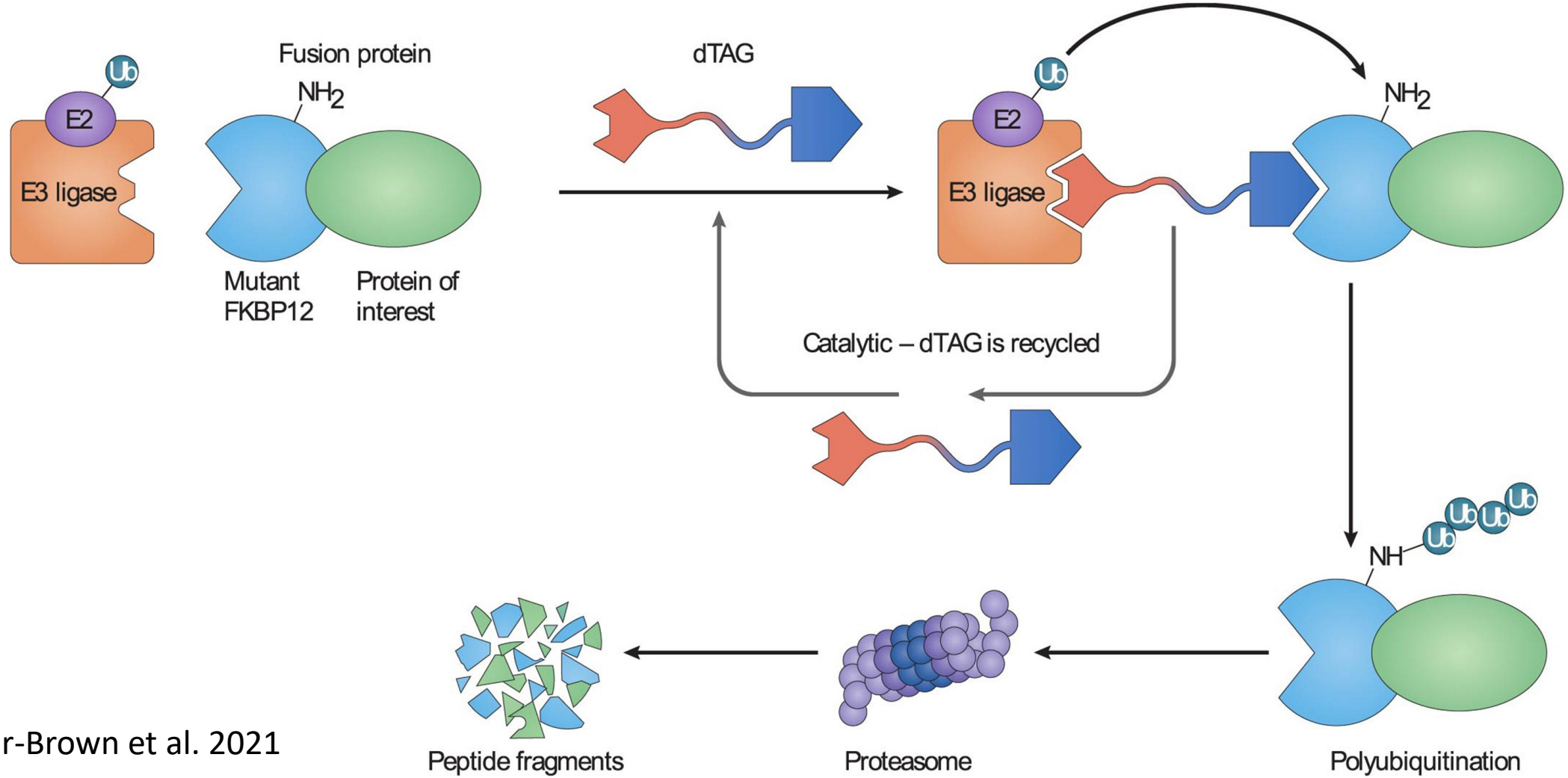
Skalska, NRMCB, 2017





Hsieh, Mol Cell, 2020

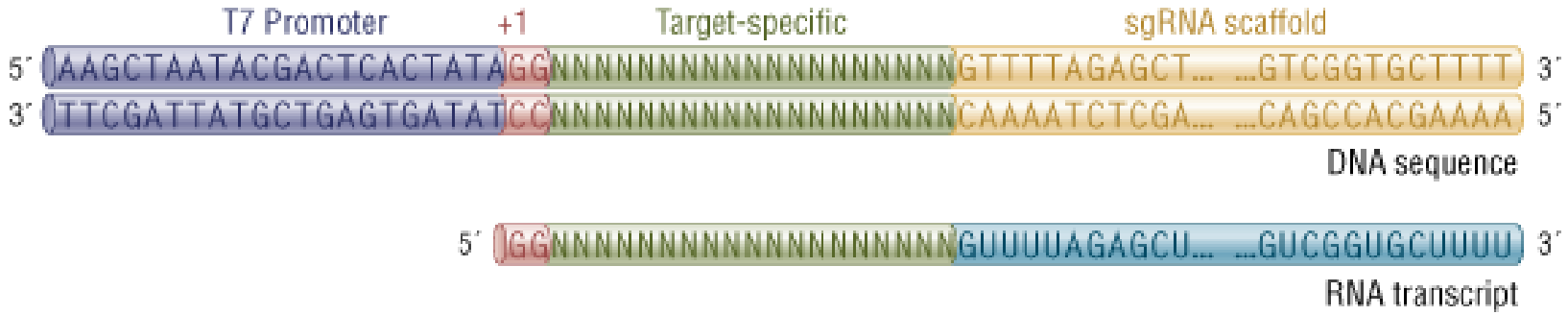
Inducible RNF20 depletion via the dTAG system



Cresser-Brown et al. 2021

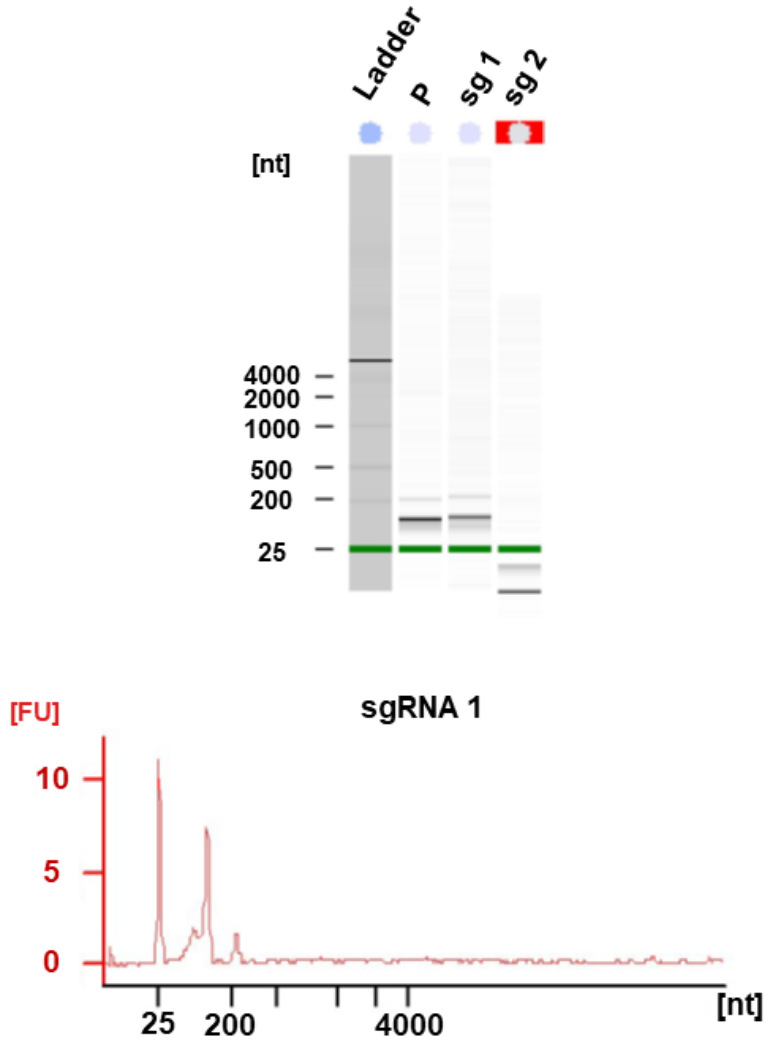
In vitro transcription of single-guide RNAs for CRISPR/Cas9 mediated knock-in of the dTAG cassette

In vitro



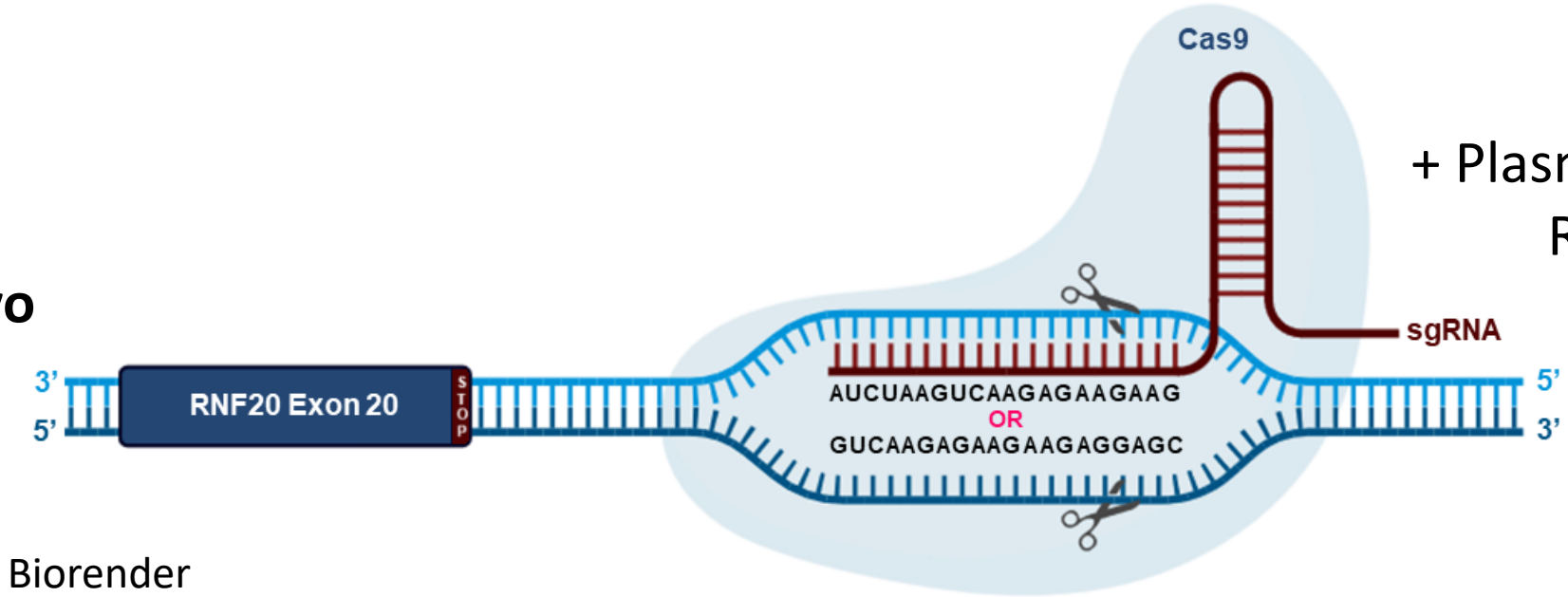
V. Theodosiou and N. Vouzounerakis

In vitro transcription of single-guide RNAs for CRISPR/Cas9 mediated knock-in of the dTAG cassette



V. Theodosiou and N. Vouzounerakis

In vivo

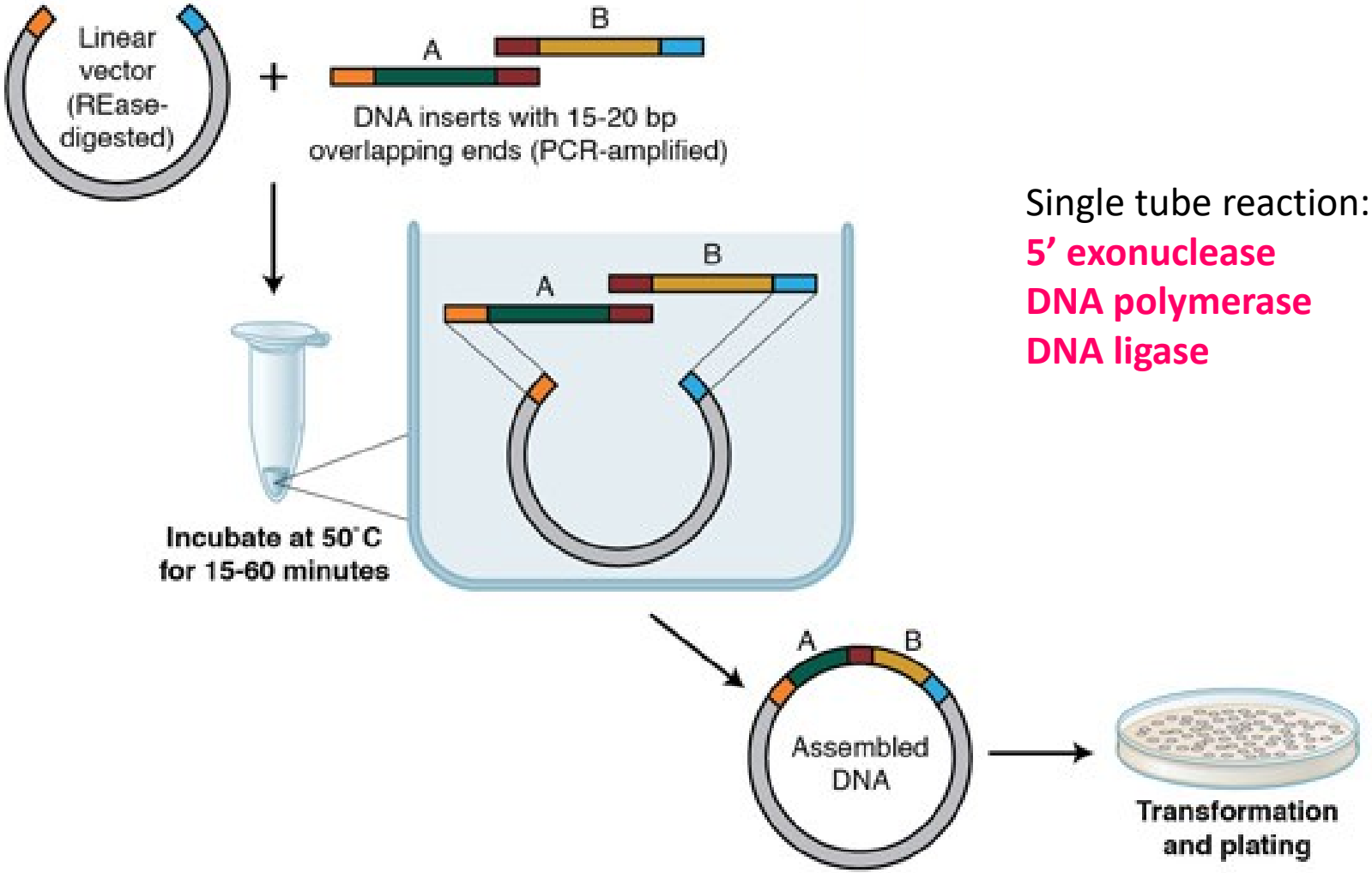


+ Plasmid for Homologous Recombination

Created with Biorender

V. Theodosiou and N. Vouzounerakis

Gibson Assembly



Single tube reaction:

5’ exonuclease

DNA polymerase

DNA ligase

Adapted from NEB

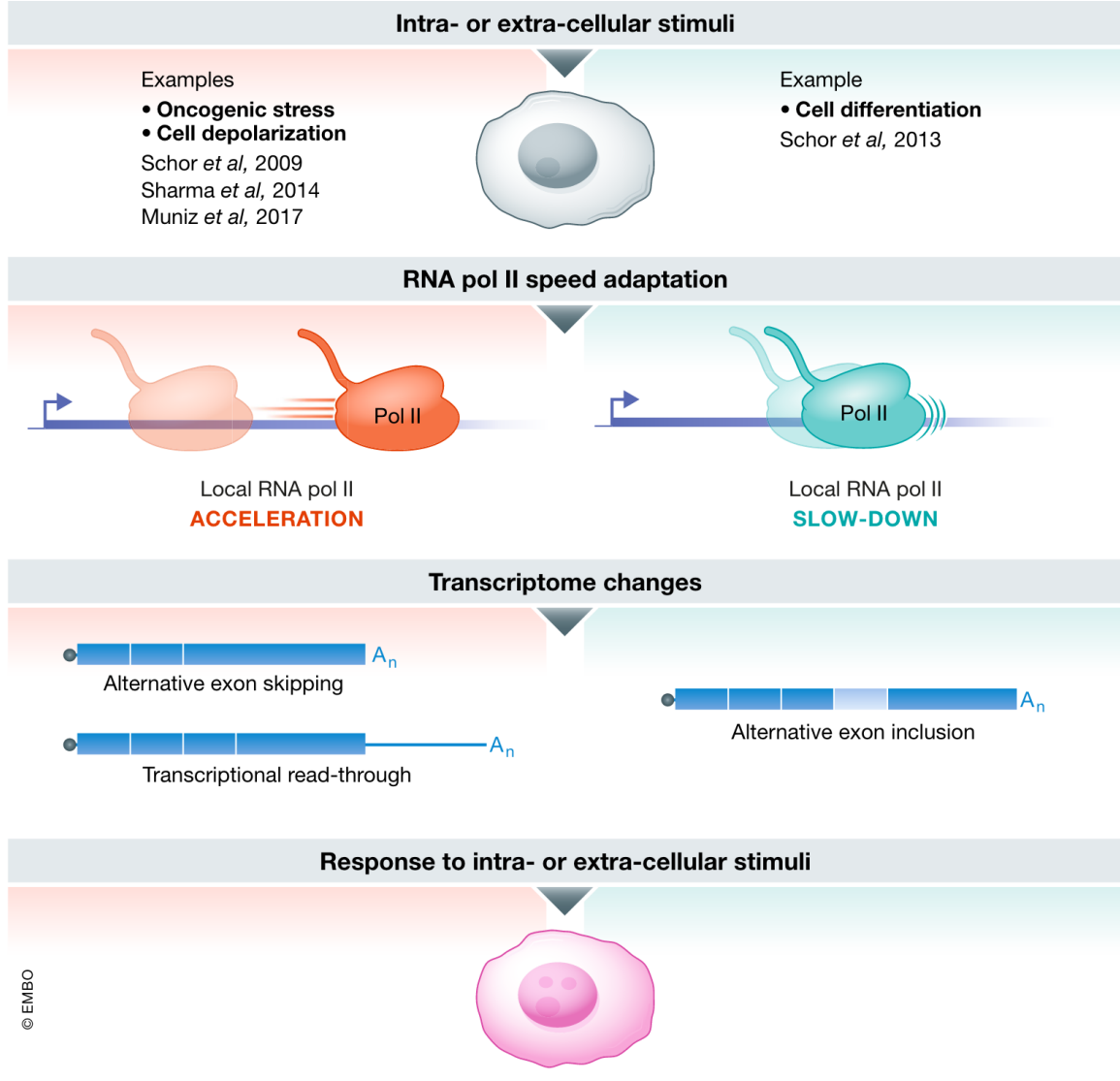


Figure 3. RNA Pol II speed is regulated in order to adapt the transcriptome composition in response to intra- or extra-cellular stimuli.

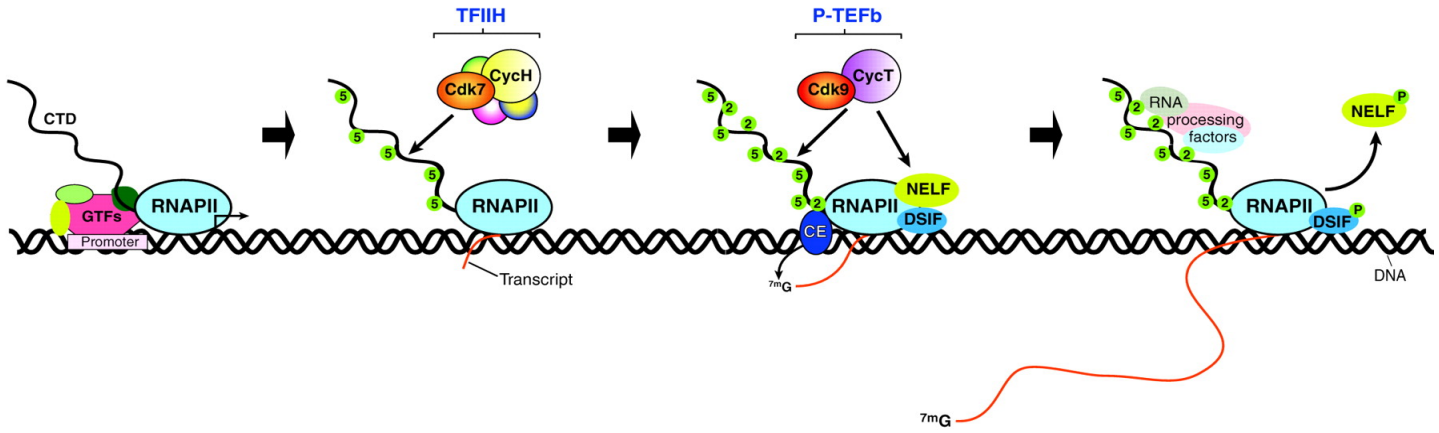
In response to intra- or extra-cellular stimuli such as oncogenic stress, cell depolarization, or cell differentiation, RNA Pol II can either accelerate or slow down locally, inducing a change in alternative splicing or the extent of read-through which could play a role in the response to stimuli.

Key Steps in RNAPII Transcription Cycle for modulating gene expression: CTD code

PIC	Initiation	PPP= Promoter-Proximal Pausing Signal-dependent transition into elongation	Elongation

RNAPII C-Terminal Domain (CTD) status

Hypo **ser5P** **ser2P**

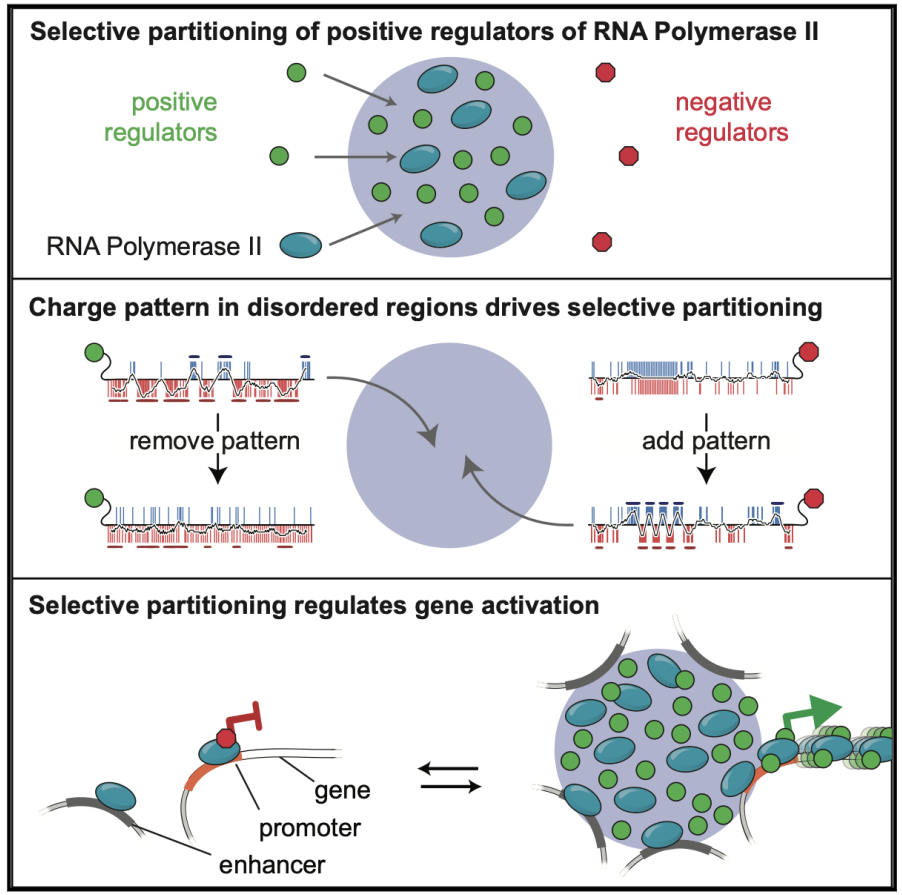


TSS **PPP**

Cell

Functional partitioning of transcriptional regulators by patterned charge blocks

Graphical abstract



Authors

Heankel Lyons, Reshma T. Veettil, Prashant Pradhan, ..., Mikayla Eppert, Robert G. Roeder, Benjamin R. Sabari

Correspondence

benjamin.sabari@utsouthwestern.edu

In brief

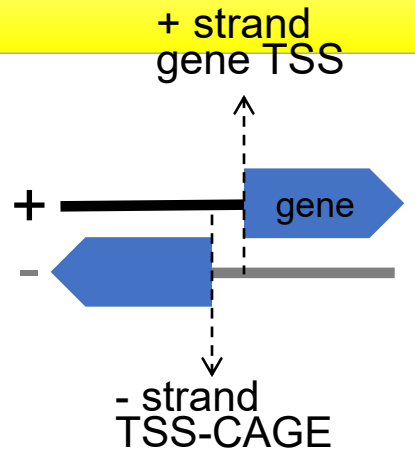
Charge patterning in disordered regions of transcriptional regulators mediates selective partitioning into MED1^{IDR} condensates for gene activation.

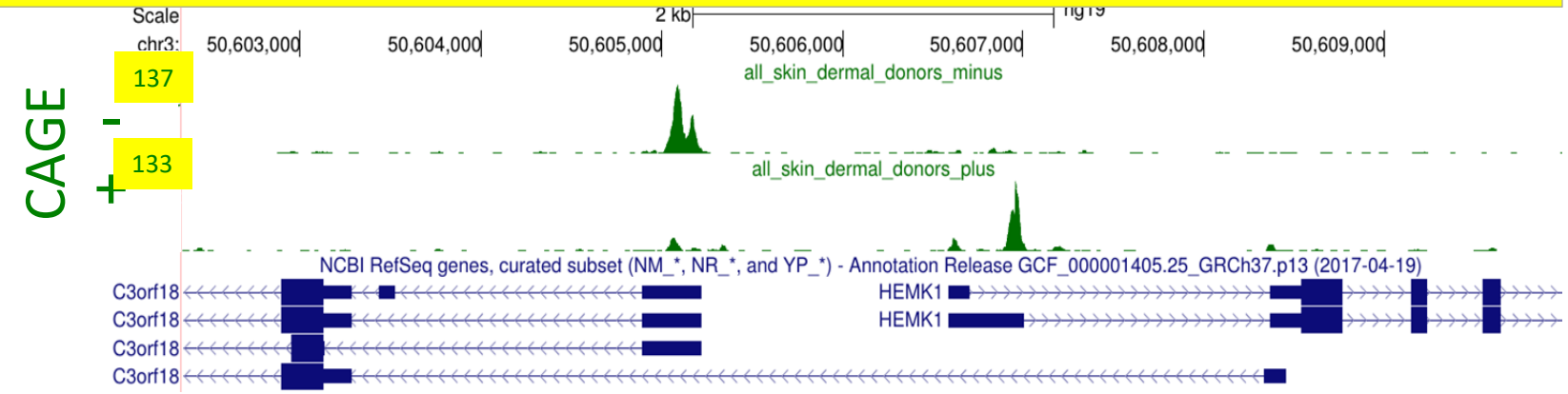
Highlights

EXTEND analysis TO ALL TRANSCRIBED UNITS regardless of transcription level and transcript stability

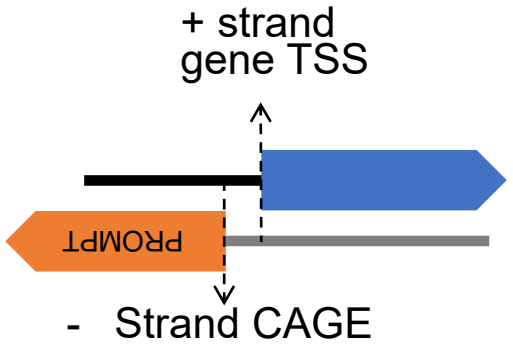
CAGE-seq data from dermal fibroblast pinpoint accurately to TSS of genes, antisense RNAs (PROMPTs) and eRNAs

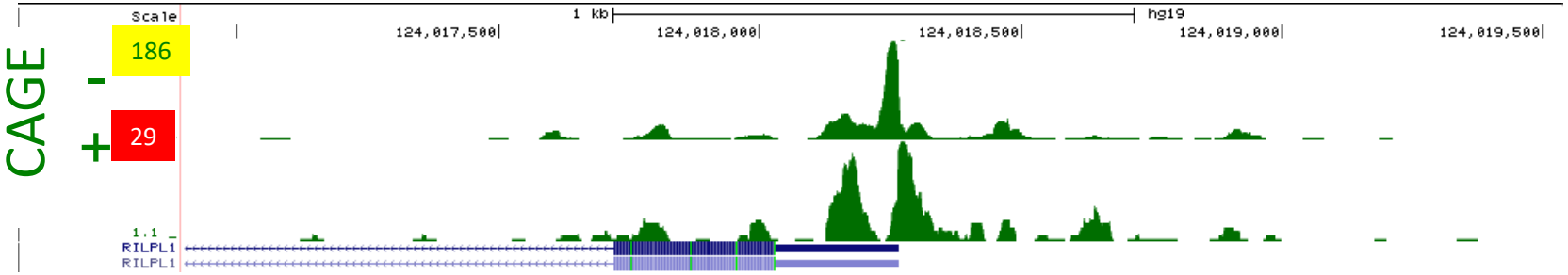
Active Promoter bidirectional



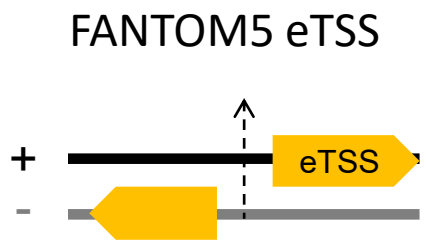


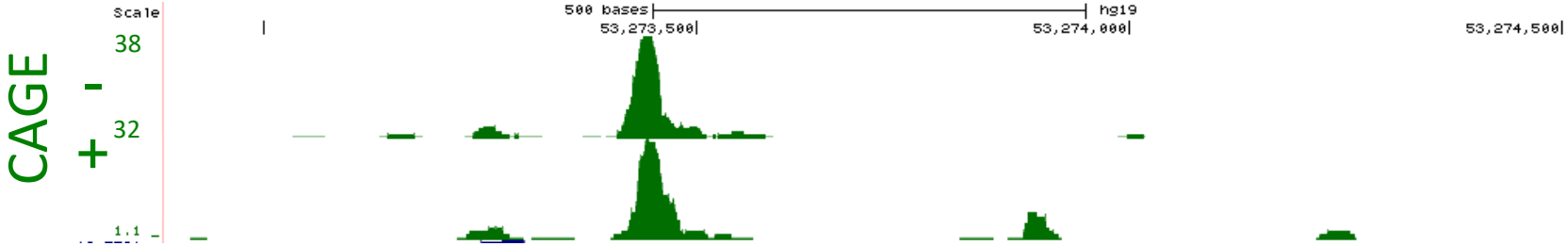
Active Promoter unidirectional



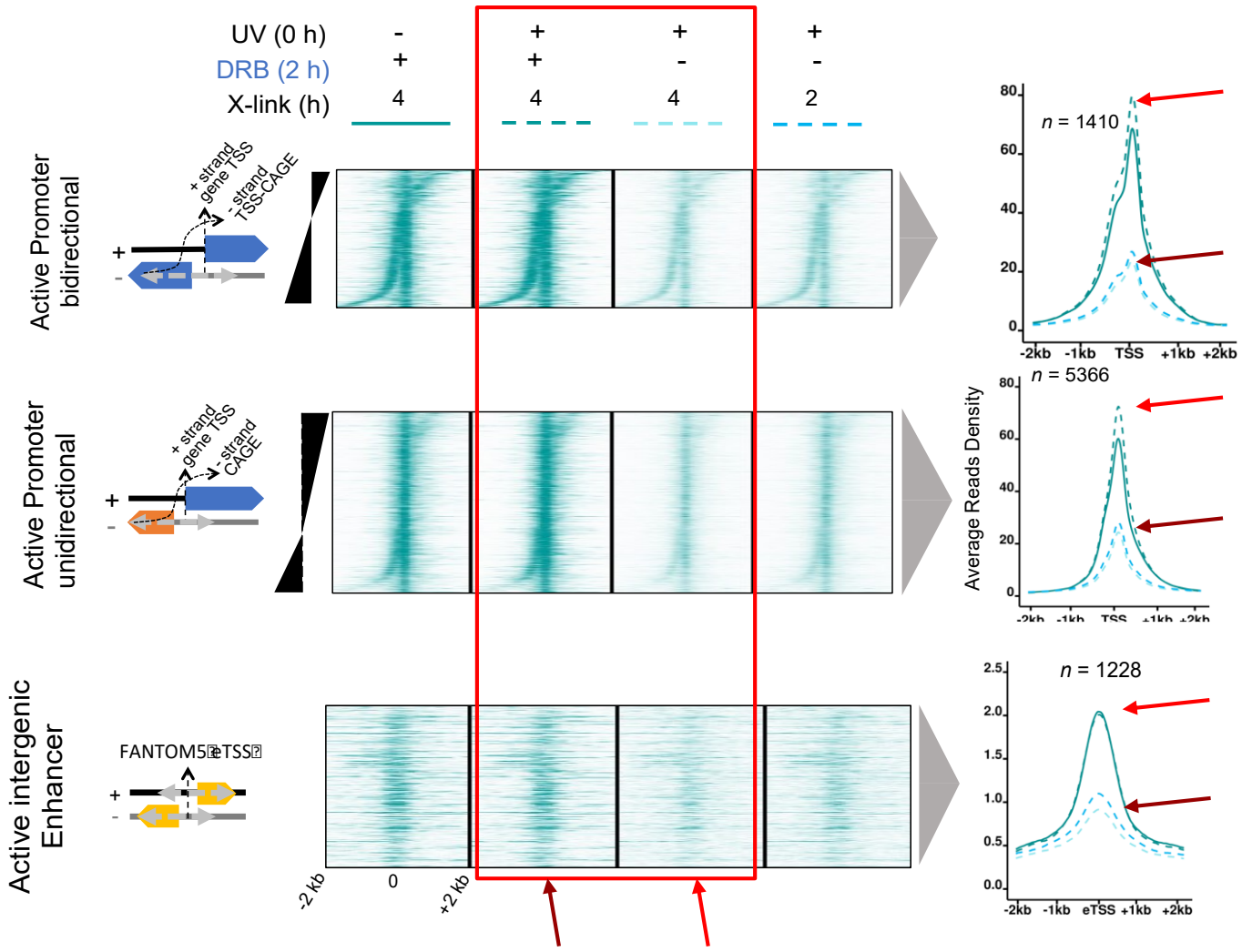


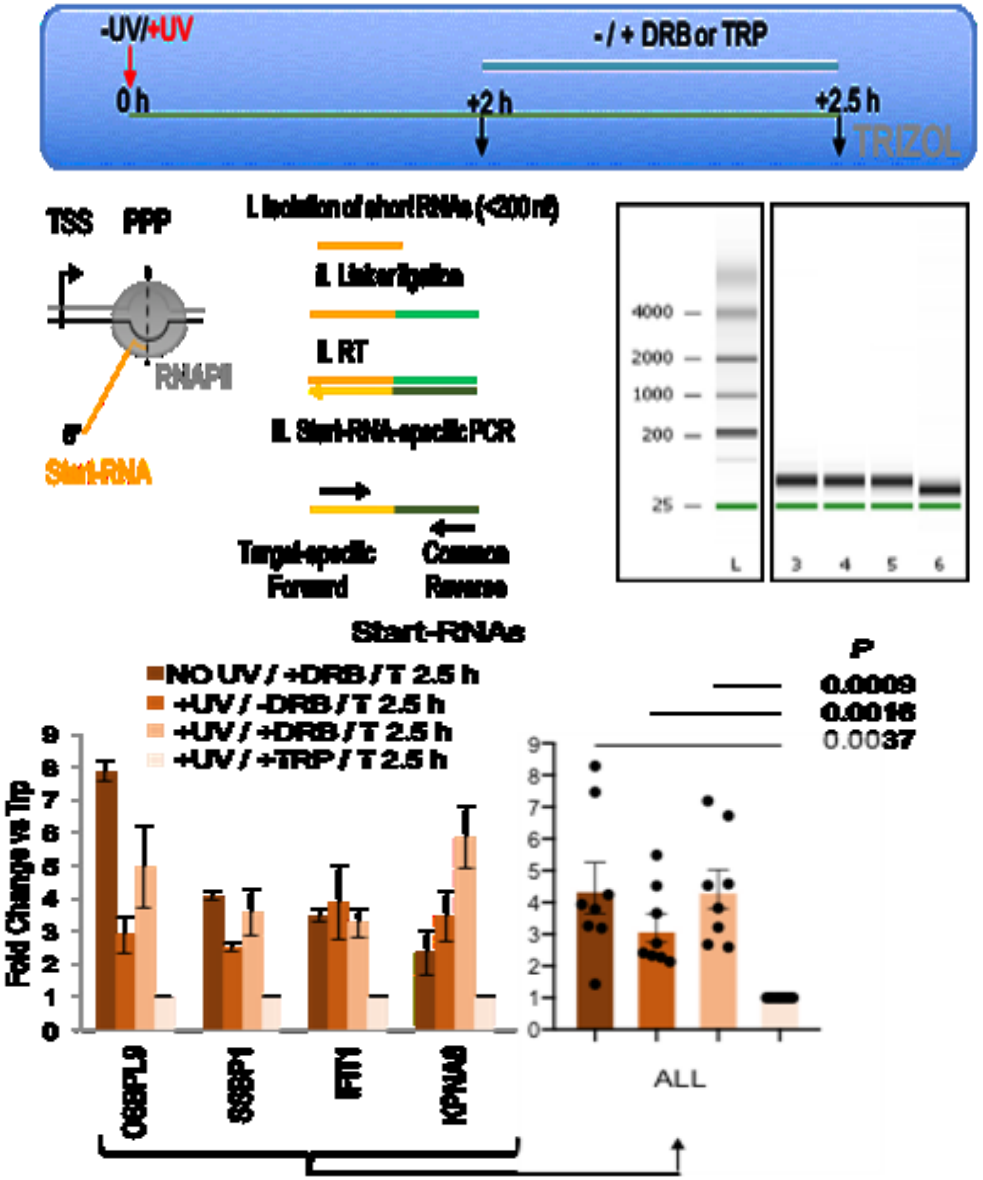
Active intergenic Enhancer





UV decreases pre-initiating pol II dwell time at ALL active TSSs
BUT continuous initiation of Pol 2 maintains high levels of start RNAs



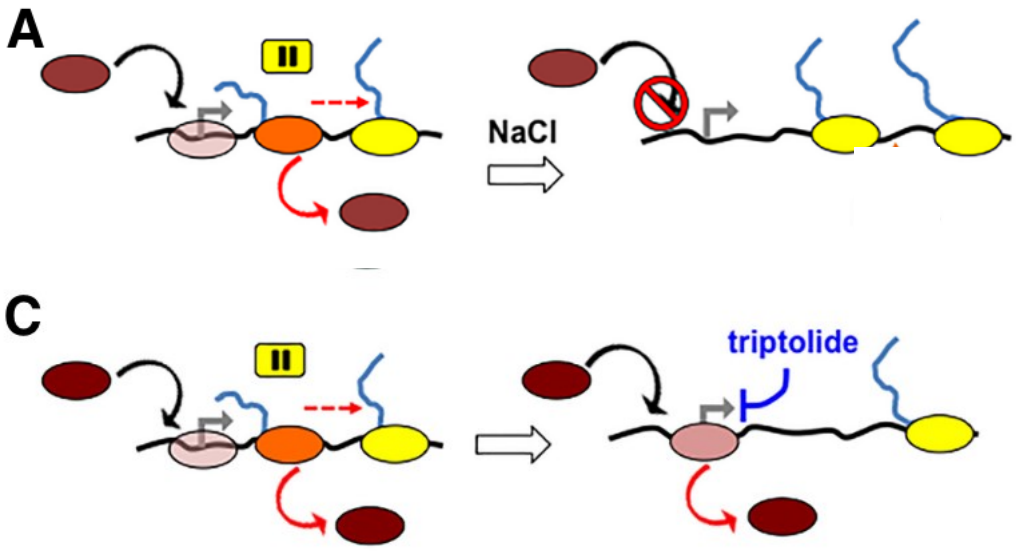


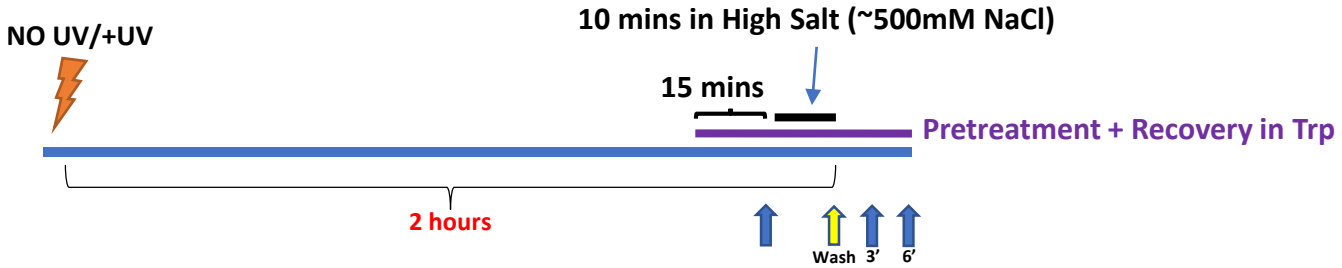
Fast transition into elongation
and NOT inhibition of initiation

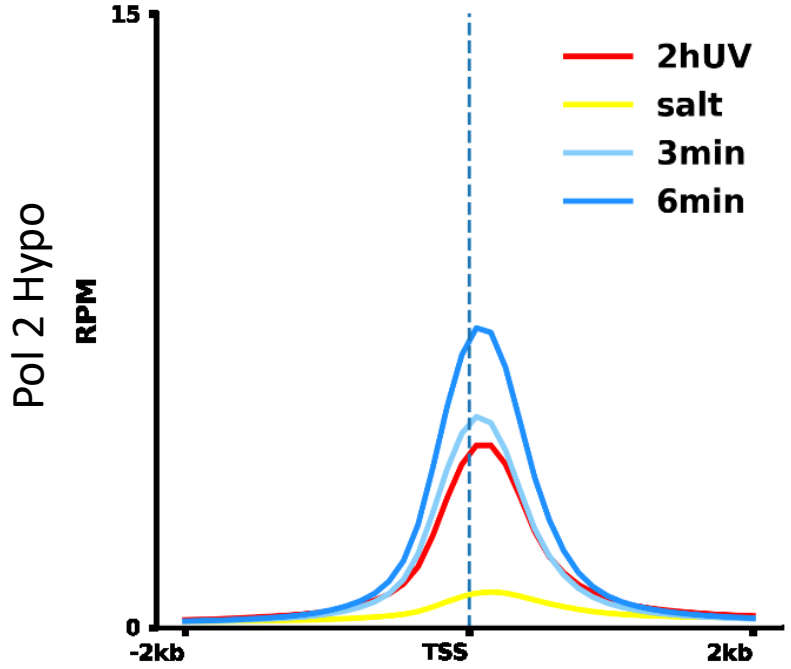
Liakos et al., 2020

Confirming recruitment of Pol 2-hypo in PIC even upon UV

Erickson et al. 2018

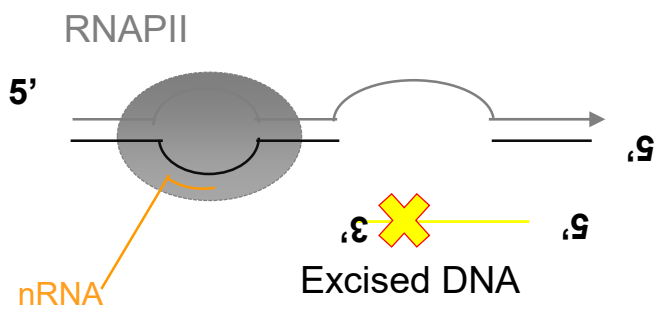






Liakos et al., in revision

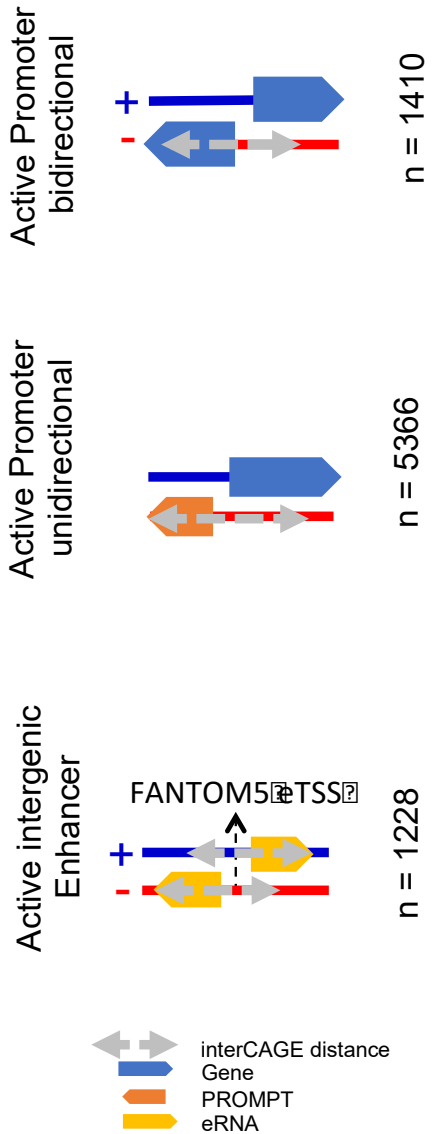
TC-NER detected at ALL active genes, PROMPTS and enhancers

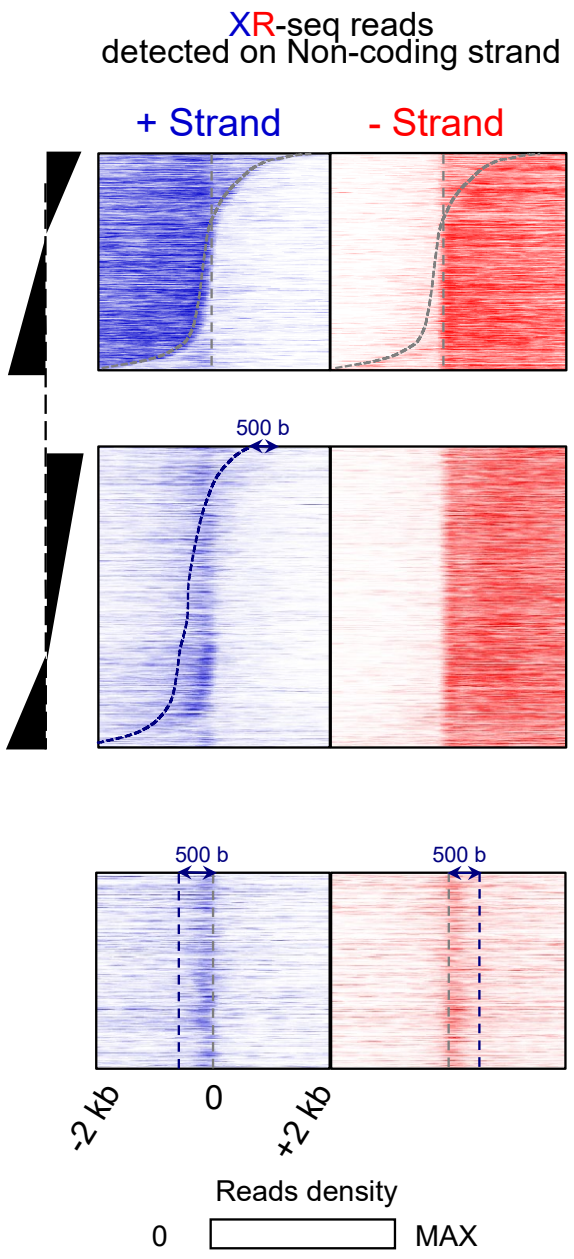


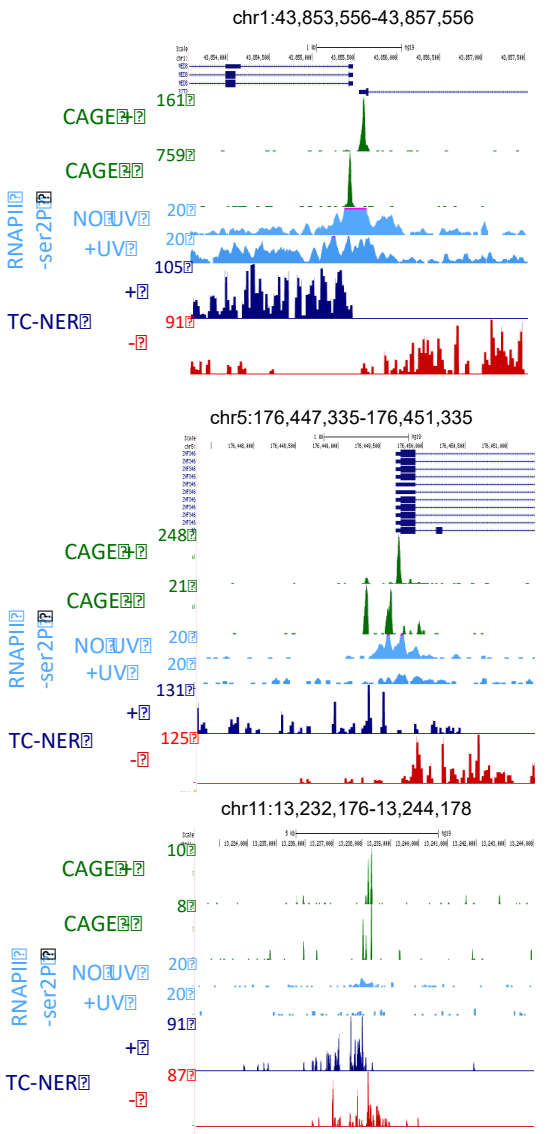
CPD

Coding / Non-template strand

Non-coding / Template strand

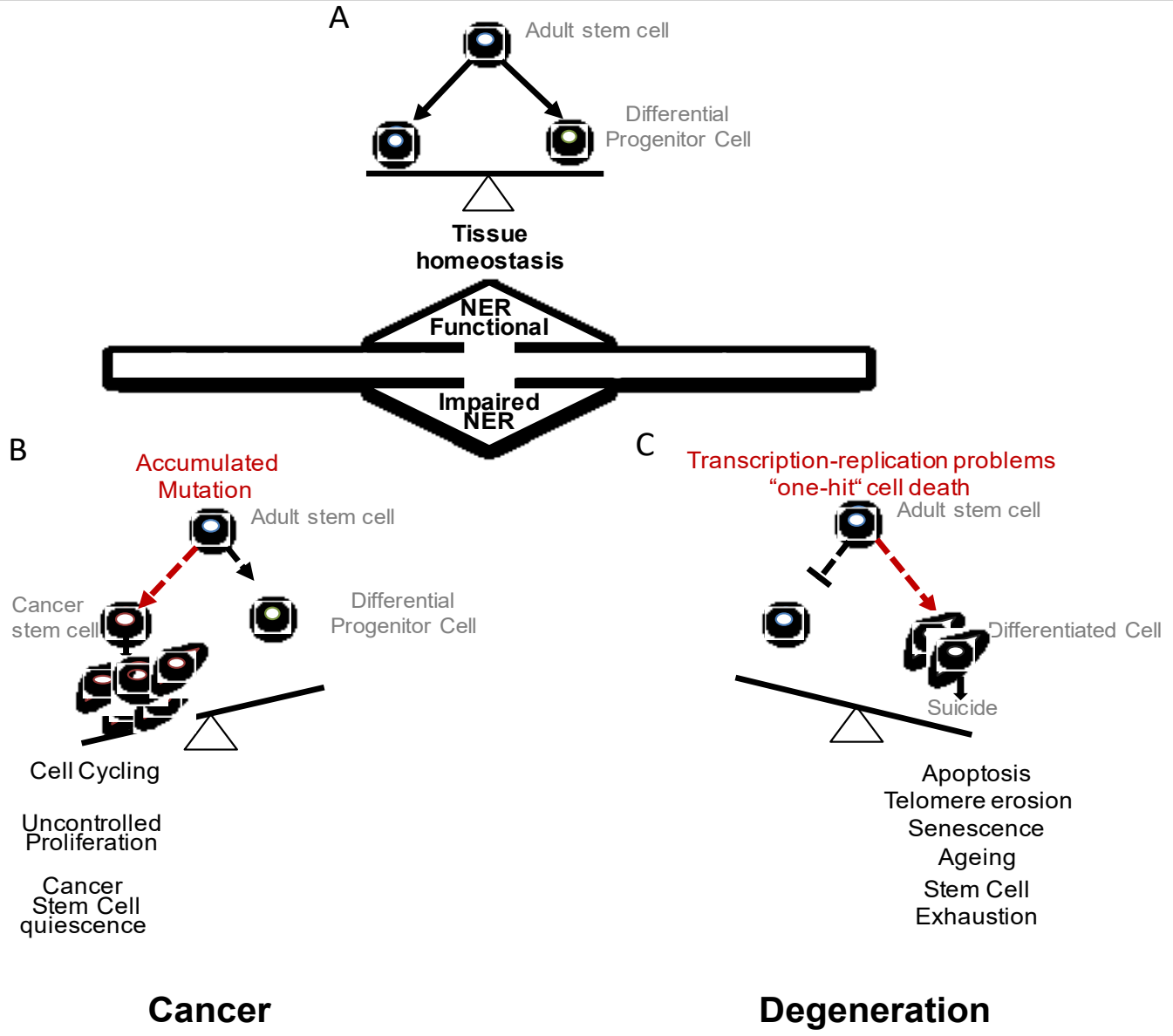






Liakos et al., 2020

Consequences of impaired/overwhelmed NER on balance between Cancer and programmed cell death depend on endogenous and exogenous parameters



Liakos et al., 2017

Curating Alexandrov et al. data for NER-specific substitutions
Filter C>T and G>T mutations on the TS and on the NTS

Pipeline construction process

Mutation name (substitution)	C > T				G > T			
Trinucleotide context	T(C)C > T(T)C				T(G)G > T(T)G			
Maximum frequency in	Melanoma				Lung Adenocarcinoma			
Reference Genome	C:G		G:C		C:G		G:C	
Cancer Genome	T:A		A:T		A:T		T:A	
Gene orientation	+	-	+	-	-	+	-	+
Mutation strand TS=Template NTS=Non-template	NTS	TS	TS	NTS	NTS	TS	TS	NTS
Select from annotated substitutions on Watson strand: TS	C > T on – genes G > A on + genes				G > T on – genes C > A on + genes			
Select from annotated substitutions on Watson strand: NTS	C > T on + genes G > A on - genes				G > T on + genes C > A on - genes			

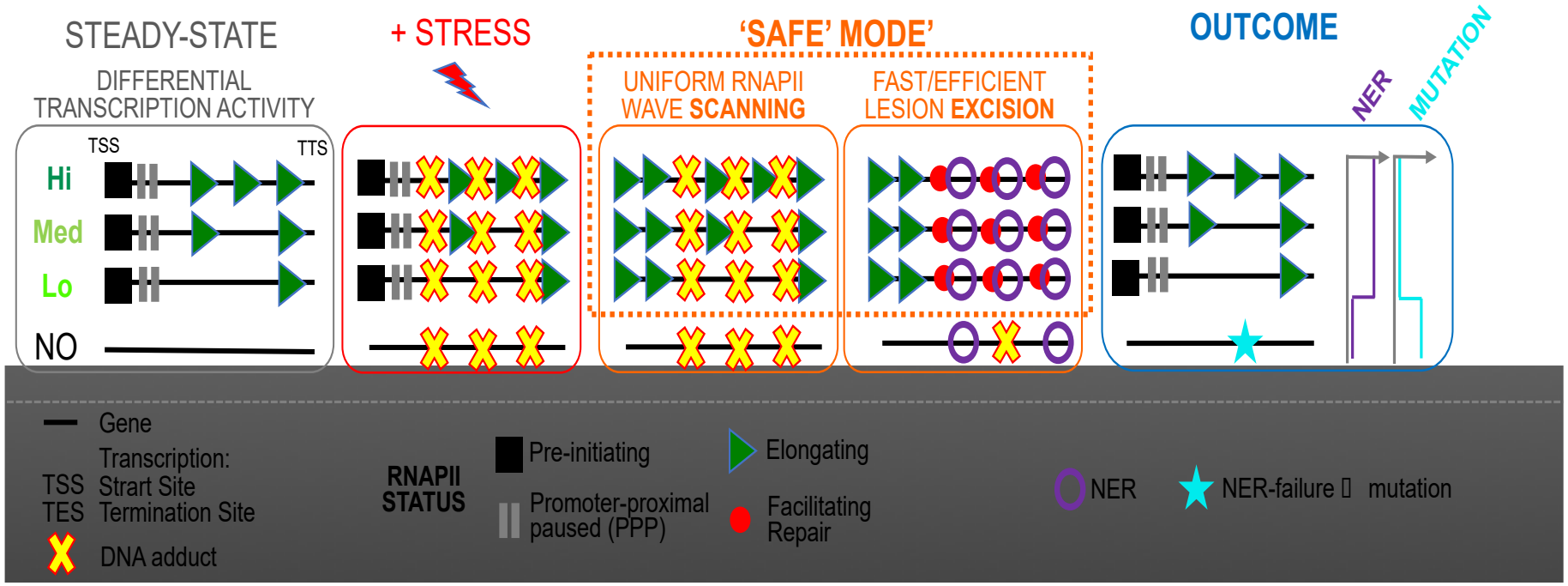
→ Make BED files for each group (x4)
In each cancer type (x2) : 8 groups

nRNA expression levels
from healthy cells (GEO) →

Map on annotated genes
ranked by
expression levels

Conclusions

Widespread RNAPII *de novo* **wave escape** and **continuous initiation**
promotes efficient genome surveillance and minimizes mutation rate uniformly in ALL transcribed regions



Lavigne et al., 2017 and Liakos et al., 2020

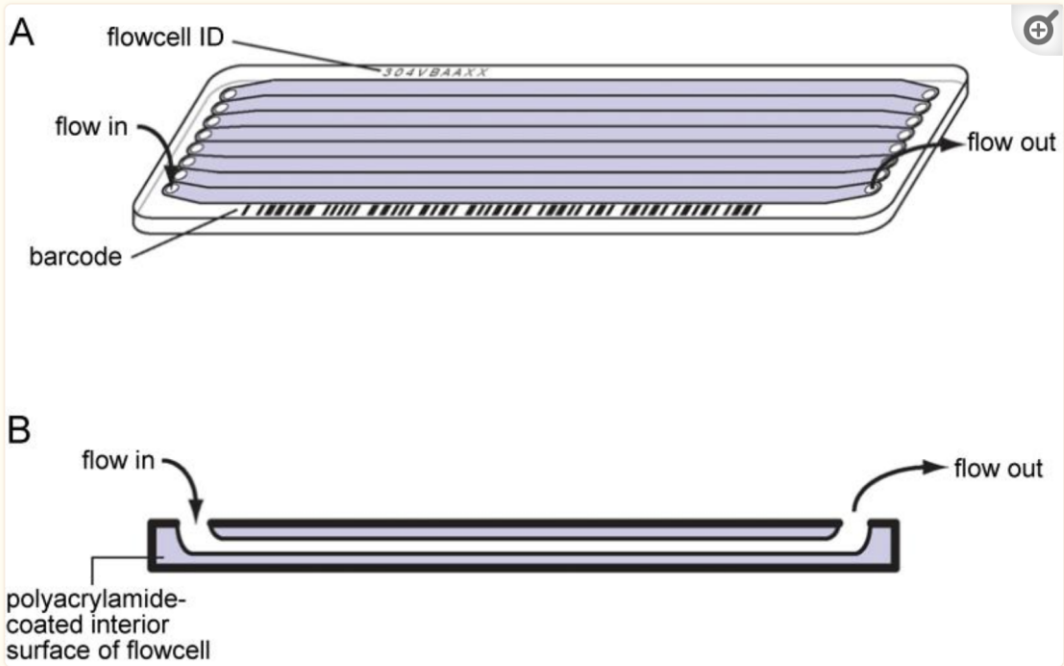


Figure 18.2.1

Illumina flowcell.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3849550/