Abstracts of papers presented
at the 2024 meeting on

TRANSPOSABLE ELEMENTS

October 15–October 19, 2024

Cold Spring Harbor Laboratory
MEETINGS & COURSES PROGRAM

Abstracts of papers presented
at the 2024 meeting on

TRANSPOSABLE ELEMENTS

October 15–October 19, 2024

Arranged by

Orsolya Barabas, *University of Geneva, Switzerland*
Cedric Feschotte, *Cornell University*
Rob Martienssen, *Cold Spring Harbor Laboratory*

Cold Spring Harbor Laboratory
MEETINGS & COURSES PROGRAM

Contributions from the following companies provide core support for the Cold Spring Harbor meetings program.

Corporate Benefactors

Estée Lauder Companies
Regeneron

Corporate Sponsors

Agilent Technologies
Biogen
Bristol-Myers Squibb Company
Calico Labs
New England Biolabs
Novartis Institutes for Biomedical Research

Cover: Photograph taken in 1947; the first transposition in maize was discovered that year.

TRANSPOSABLE ELEMENTS
Tuesday, October 15 – Saturday, October 19, 2024

Tuesday	7:30 pm – 8:15 pm	**Keynote Speaker I**
Tuesday	8:30 pm – 9:30 pm	**1** Impacts on the Host I: Early Development
Wednesday	9:00 am – 12:00 pm	**2** Transposon Life Cycles I: Mechanisms of Transposition
Wednesday	2:00 pm – 5:00 pm	**3** Impacts on the Host II: From Ageing to Regeneration
Wednesday	5:00 pm	*Wine & Cheese Party*
Wednesday	7:30 pm – 10:30 pm	**Poster Session I**
Thursday	9:00 am – 12:00 pm	**4** Transposon Control I: Host Factors
Thursday	2:00 pm – 2:45 pm	**Keynote Speaker II**
Thursday	2:45 pm – 5:00 pm	**5** Transposon Life Cycles II: Site Specificity
Thursday	7:30 pm – 10:30 pm	**Poster Session II**
Friday	9:00 am – 12:00 pm	**6** Transposon Control II: Germline vs. Soma
Friday	2:00 pm – 5:00 pm	**7** Transposon Life Cycles III: DNA Rearrangements
Friday	6:00 pm 7:00 pm	*Concert* *Cocktails and Banquet*
Saturday	9:00 am – 12:00 pm	**8** Variety and Evolution of Transposable Elements

All times shown are US Eastern: Time Zone Converter

Mealtimes at Blackford Hall are as follows:
Breakfast 7:30 am-9:00 am
Lunch 11:30 am-1:30 pm
Dinner 5:30 pm-7:00 pm

Bar is open from 5:00 pm until late

Cold Spring Harbor Laboratory is committed to maintaining a safe and respectful environment for all meeting attendees, and does not permit or tolerate discrimination or harassment in any form. By participating in this meeting, you agree to abide by the Code of Conduct.

For further details as well as Definitions and Examples and how to Report Violations, please see the back of this book.

Abstracts are the responsibility of the author(s) and publication of an abstract does not imply endorsement by Cold Spring Harbor Laboratory of the studies reported in the abstract.

These abstracts should not be cited in bibliographies. Material herein should be treated as personal communications and should be cited as such only with the consent of the author(s).

Please note that photography or video/audio recording of oral presentations or individual posters is strictly prohibited except with the advance permission of the author(s), the organizers, and Cold Spring Harbor Laboratory.

Any discussion via social media platforms of material presented at this meeting requires explicit permission from the presenting author(s).

Printed on 100% recycled paper.

PROGRAM

TUESDAY, October 15—7:30 PM

KEYNOTE SPEAKER I

Introduction by: Cedric Feschotte, Cornell University

Transposable elements in cellular plasticity and reprogramming
Maria-Elena Torres-Padilla.
Presenter affiliation: Institute of Epigenetics & Stem Cells, Munich, Germany. 1

TUESDAY, October 15—8:30 PM

SESSION 1 IMPACTS ON THE HOST I: EARLY DEVELOPMENT

Chairperson: **Andrew Modzelewski,** University of Pennsylvania, Philadelphia

Novel Insights into ERV regulation—Dynamic interplay between heterochromatin and transcriptional activators
Angela Russo.
Presenter affiliation: Biomedical Center (BMC), Faculty of Medicine, Ludwig-Maximilians-University (LMU), Munich, Germany. 2

Activities of the domesticated retrotransposon PEG10 and the closely related gene RTL8
Autumn M. Matthews, Will Campodonico, Phuoc Huynh, Holly H. Black, Alexandra M. Whiteley.
Presenter affiliation: University of Colorado Boulder, Boulder, Colorado. 3

A DNA demethylase antagonizes parent-of-origin specific methylation asymmetry in Arabidopsis endosperm
Elizabeth A. Hemenway, Mary Gehring.
Presenter affiliation: Whitehead Institute, Cambridge, Massachusetts; MIT, Cambridge, Massachusetts. 4

SESSION 2 TRANSPOSON LIFE CYCLES I: MECHANISMS OF TRANSPOSITION

Chairperson: **John Moran,** University of Michigan Medical School, Ann Arbor

< Back

Evolution of immune splicing regulation mediated by transposable elements
Edward B. Chuong.
Presenter affiliation: University of Colorado Boulder, Boulder, Colorado. 16

Transposon-supported regeneration in planarians
Hae-Lim Lee, Axel Poulet, Sudheesh Allikka Parambil, Josien C. van Wolfswinkel.
Presenter affiliation: Yale University, New Haven, Connecticut. 17

Long-read sequencing pinpoints the genomic sources and regulatory determinants of endogenous retrovirus expression in cancer
Justin S. Becker, Daniel Fridman, Kevin Dong, Gary Sun, Maheshwaran Natarajan, Bradley E. Bernstein.
Presenter affiliation: Dana-Farber Cancer Institute, Boston, Massachusetts; Massachusetts General Hospital Cancer Center, Boston, Massachusetts; Broad Institute of MIT and Harvard, Cambridge, Massachusetts. 37

The Yin Yang of retrotransposons activity in the soma—From harmful to resilient RNAs in aging and tissue regeneration
Valerio Orlando.
Presenter affiliation: KAUST, Thuwal, Saudi Arabia. 19

Regulation of active LINE1 elements extends lifespan and healthspan in mice
John C. Martinez, Matthew Simon, Francesco Morandini, Lucinda Fitzgibbons, Sung Jae Bae, Natasha Sieczkiewicz, Seyed Ali Biashad, Maxfield Kelsey, Michael Meadow, John Sedivy, Andrei Seluanov, Vera Gorbunova.
Presenter affiliation: University of Rochester, Rochester, New York. 20

WEDNESDAY, October 16—5:00 PM

Wine and Cheese Party

WEDNESDAY, October 16—7:30 PM

POSTER SESSION I

See p. xvi for List of Posters

SESSION 4 TRANSPOSON CONTROL I: HOST FACTORS

Chairperson: **Michael Imbeault,** University of Cambridge, United Kingdom

The regulatory potential of transposable elements in maize
Kerry L. Bubb, Morgan O. Hamm, Joseph K. Min, Bryan Ramirez-Corona, Nicholas A. Mueth, Jane Ranchalis, Mitchell R. Vollger, Cole Trapnell, Josh T. Cuperus, Christine Queitsch, Andrew B. Stergachis.
Presenter affiliation: University of Washington, Seattle, Washington. 26

Transposon amplification in Arabidopsis expressing the oncohistone H3.1K27M
Yannick Jacob.
Presenter affiliation: Yale University, New Haven, Connecticut. 27

The genetic basis of natural DNA methylation variation on transposable elements in *Arabidopsis thaliana*
Pierre Bourguet, Magnus Nordborg, Frédéric Berger, Eriko Sasaki.
Presenter affiliation: Kyushu University, Fukuoka, Japan. 28

THURSDAY, October 17—2:00 PM

KEYNOTE SPEAKER II

Introduction by: Rob Martienssen, Cold Spring Harbor Laboratory / HHMI

Centrophilic retrotransposon integration via CENH3 chromatin in Arabidopsis
Tetsuji Kakutani.
Presenter affiliation: University of Tokyo, Tokyo, Japan. 29

THURSDAY, October 17—2:45 PM

SESSION 5 TRANSPOSON LIFE CYCLES II: SITE SPECIFICITY

Chairperson: **Orsolya Barabas,** University of Geneva, Switzerland

Who and where? Uncovering active transposable elements and their contextual mobilization preferences in Arabidopsis
Leandro Quadrana.
Presenter affiliation: Institute of Plant Science Paris-Saclay, Gif-sur-Yvette, France. 30

THURSDAY, October 17—7:30 PM

POSTER SESSION II

See p. xxvi for List of Posters

SESSION 6 TRANSPOSON CONTROL II: GERMLINE vs. SOMA

Chairperson: **Laure Teysset,** Sorbonne Université, CNRS, Paris, France

FRIDAY, October 18—2:00 PM

SESSION 7 TRANSPOSON LIFE CYCLES III: DNA REARRANGEMENTS

Chairperson: **Nancy Craig,** SalioGen Therapeutics, Lexington, Massachusetts

Antagonistic conflict between transposon-encoded introns and guide RNAs
Rimante Zedaveinyte, Chance Meers, Hoang C. Le, Edan E. Mortman, Stephen Tang, George D. Lampe, Sanjana R. Pesari, Diego R. Gelsinger, Tanner Wiegand, Younggi D. Moon, Samuel H. Sternberg.
Presenter affiliation: Columbia University, New York, New York. 47

TE mobility differs in serial isolates of *Cryptococcus neoformans* from recurrent human infections
Anna Mackey, Vesper Fraunfelter, Samantha Shaltz, John Perfect, Sue Jinks-Robertson, Asiya Gusa.
Presenter affiliation: Duke University, Durham, North Carolina. 48

Large serine integrases—How do they know which way to go?
Heewhan Shin, Alexandria Holland, Abdulrazak Alsaleh, Alyssa D. Retiz, Ying Z. Pigli, Oluwateniola T. Taiwo-Aiyerin, Tania Peña Reyes, Adebayo J. Bello, Femi J. Olorunniji, Phoebe A. Rice.
Presenter affiliation: The University of Chicago, Chicago, Illinois. 49

Activity of the mammalian DNA transposon *piggyBat* from *Myotis lucifugus* is inhibited by its own transposon ends
Alison B. Hickman, Laurie Lannes, Christopher M. Furman, Christina Hong, Lidiya Franklin, Fred Dyda.
Presenter affiliation: National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland. 50

Targeted genomic integration of a large DNA in the absence of DNA double strand breaks by a programmable mammalian transposase—Insertion of a therapeutic CFTR gene into Intron 1 of the CFTR gene in primary human cells
Nancy Craig, Subhrangshu Guhathakurta, Jaclyn M Kubala, Sharmishtha Musalgaonkar, Daniel Pers, Rupesh Naidu, Esther Epum, Lillian Campos, Sarah Slauson, Joseph J Higgins, Joseph J Senn.
Presenter affiliation: SalioGen Therapeutics, Lexington, Massachusetts. 51

FRIDAY, October 18—6:00 PM

CONCERT
Grace Auditorium

Teddy Siegel, *soprano*
Toby Bradford, *tenor*
Maxim Lando, *piano*

FRIDAY, October 18—7:00 PM

COCKTAILS and BANQUET

SATURDAY, October 19—9:00 AM

SESSION 8 VARIETY AND EVOLUTION OF TRANSPOSABLE ELEMENTS

Chairperson: **Irina Arkhipova,** Marine Biological Laboratory, Woods Hole, Massachusetts

Comparative genomics, phylogenetic analysis, and modeling of transposable element dynamics across *Caenorhabditis* Nematode Genomes
Victoria Eggers, Rohit Kapila, Gabriel Blanco, Sayran Saber, Jason Pienaar, Janna Fierst.
Presenter affiliation: Florida International University, Miami, Florida. 51

Towards an unbiased characterization of genetic diversity
Anna A. Igolkina, Sebastian Vorbrugg, Fernando A. Rabanal, Hai-Jun Liu, Haim Ashkenazy, Aleksandra E. Kornienko, Detlef Weigel, Magnus Nordborg.
Presenter affiliation: Austrian Academy of Sciences, Vienna, Austria. 52

Transgenerational epigenetic variation at TE sequences in Arabidopsis—Molecular determinants and impact in nature
Pierre Baduel, Louna De Oliveira, Erwann Caillieux, Grégoire Bohl-Viallefond, Mounia El Messaoudi, Ciana Xu, Matteo Barois, Vipin Singh, Alexis Sarazin, Martine Boccara, Elodie Gilbault, Antoine de France, Leandro Quadrana, Olivier Loudet, Vincent Colot.
Presenter affiliation: ENS/PSL, PARIS, France. 53

Junk DNA? Megabase-scale heterochromatin removal allows functional interrogation of *Arabidopsis* pericentromeric regions
Valentin Joly, Máximo Kesselhaut, Sourav Roy, Yannick Jacob.
Presenter affiliation: Yale University, New Haven, Connecticut. 54

Phased T2T ape genomes show differential TE insertion rates
Mark Loftus, Ashley E. Kirby, Parithi Balanchandran, Jessica M. Storer, Weichen Zhou, The T2T Ape Consortium, Ryan E. Mills, Evan E. Eichler, Rachel J. O'Neill, Christine R. Beck, Miriam K. Konkel.
Presenter affiliation: Clemson University, Clemson, South Carolina. 55

< Back

POSTER SESSION II

< Back

TRANSPOSABLE ELEMENTS IN CELLULAR PLASTICITY AND REPROGRAMMING

Maria-Elena Torres-Padilla

Institute of Epigenetics & Stem Cells, Helmholtz Munich, Munich, Germany

Transposable elements and their remnants occupy more than half of mammalian genomes. The regulatory circuitry of cell-specific transcriptional programmes is thought to be influenced by transposable elements (TEs). In addition, the expression of TEs themselves plays a role in shaping cell fate and identity during development. In particular, the earliest developmental stages in mammals are characterised by a robust transcriptional activation of TEs. We have investigated their role, evolutionary and age expression as well as their transcriptional activators during early development. We will discuss these new data within a framework of co-option of TEs by the host for fundamental physiological processes such as development and reprogramming.

NOVEL INSIGHTS INTO ERV REGULATION: DYNAMIC INTERPLAY BETWEEN HETEROCHROMATIN AND TRANSCRIPTIONAL ACTIVATORS.

Angela Russo

Biomedical Center (BMC), Faculty of Medicine, Ludwig-Maximilians-University (LMU), Division of Molecular Biology, Munich, Germany

Endogenous retroviruses (ERVs) are widespread across the genome and can affect gene regulation by acting as enhancers or alternative promoters under physiological conditions. However, any aberrant activity must be prevented to maintain cell type specific expression patterns. Therefore, ERV activity is governed by mechanisms that ensure their cell type specific activation or repression.
A major mechanism for ERV repression is heterochromatin formation, which is thought to inhibit the binding of transcription factors (TFs) on ERVs. The main heterochromatin features on ERVs are DNA methylation and the repressive histone modifications H3K9me3 and H4K20me3.
ERV activity is cell type dependent and requires conversion of active and inactive chromatin states during developmental transitions. We investigated changes in heterochromatin features by profiling chromatin accessibility, active histone modifications, and repressive histone modifications in mouse embryonic stem cells (mESCs) and extraembryonic endoderm cells (XEN). Surprisingly, we detected ERVs that gain H3K27Ac, a hallmark of active enhancers, on H3K9me3/H4K20me3-marked ERVs in the transition to XEN cells. This finding suggests that activating transcription factors may counteract heterochromatin in this context. Motif analysis on ERV sequences exhibiting these distinctive chromatin features revealed p53 binding sites in the majority of these ERVs. ChIPseq analysis of p53 confirmed binding to these elements, demonstrating that p53 can bind ERV heterochromatin.
To verify that heterochromatin is necessary to counteract activation mechanisms of p53 bound ERVs in XEN cells, we impaired heterochromatin establishment by mutating Daxx, an important component of the silencing machinery. Consistently, in Daxx ko XEN cells we found a global reduction in heterochromatin markers. Interestingly, we only detected increased H3K27ac and chromatin accessibility on elements that already showed low levels of H3K27ac in wild type cells. This suggests that heterochromatin is dispensable on regions with no transcription factor binding, but needed to counteract transcription factor activity on p53 bound ERVs. RNAseq analysis revealed higher transcription of genes neighboring these elements in Daxx ko XEN cells, indicating that p53 bound ERVs gain aberrant enhancer activity in absence of heterochromatin.
In summary, our data demonstrate that heterochromatin does not generally prevent binding of TFs, but that transcription factor activity is restricted by currently unknown mechanisms, to prevent aberrant TF activity on non-physiological binding sites.

ACTIVITIES OF THE DOMESTICATED RETROTRANSPOSON PEG10 AND THE CLOSELY RELATED GENE RTL8

Autumn M Matthews, Will Campodonico, Phuoc Huynh, Holly H Black, Alexandra M Whiteley

University of Colorado Boulder, Biochemistry, Boulder, CO

PEG10 is a domesticated retroelement expressed in humans and present in all mammals that encodes a retrotransposon-like gag-pol gene. Animal studies have shown a critical role for PEG10 in placental development, but its mechanism of activity is both complex and unclear. PEG10 is one of very few human genes with a programmed ribosomal frameshifting site which allows for production of gag and gag-pol proteins from the same mRNA transcript. We have seen a dramatic increase in PEG10 gag-pol protein levels in the disease Amyotrophic Lateral Sclerosis (ALS), especially in ALS cases caused by UBQLN2 mutation, implying a balance between adaptive roles in the placenta and pathological roles in the nervous system. Here, I will discuss two recent investigations into PEG10 activity. First, we demonstrate retrotransposon-like activities of the gag-pol form of PEG10 that result in changes to neuronal gene expression reminiscent of neurodegeneration. PEG10 is capable of retroelement-like self-cleavage to generate a liberated, RNA-binding 'nucleocapsid' fragment which uniquely shuttles to the nucleus. Recombinant NC protein binds to RNA with specific sequence preferences, and overexpression of just the nucleocapsid fragment alters expression of neuronal genes, including DCLK1, which regulates dendritic spines and is altered in human ALS. Second, we explore the regulation of PEG10-derived virus-like particles (VLPs). We find that multiple human cell lines spontaneously release PEG10 VLPs, and that particle release is negatively influenced by co-expression of a closely related Mart family member, RTL8. Like PEG10, RTL8 is a gag-like gene, but is only 113 amino acids in length and does not contain a full capsid domain or zinc finger. These findings implicate the Mart family member RTL8 in the inhibition of PEG10 VLP formation or release, and may have implications for human health and engineering of VLPs for therapeutic delivery.

A DNA DEMETHYLASE ANTAGONIZES PARENT-OF-ORIGIN SPECIFIC METHYLATION ASYMMETRY IN ARABIDOPSIS ENDOSPERM

Elizabeth A Hemenway[1,2], Mary Gehring[1,2]

[1]Whitehead Institute, Cambridge, MA, [2]MIT, Biology, Cambridge, MA

The endosperm is a critical seed tissue with a unique genetic and epigenetic landscape, in part shaped by active removal of 5-methylcytosine (DNA methylation) prior to fertilization. Endosperm is characterized by transposable element hypomethylation and expression. In *Arabidopsis thaliana* there are four DNA demethylases: DME, ROS1, DML2, and DML3. DME is active in the central cell of the female gametophyte prior to fertilization, and is required for gene imprinting and endosperm viability. DME thus promotes epigenetic asymmetry between maternal and paternal genomes in the endosperm. *ROS1* is not an essential gene, but is the most transcriptionally abundant DNA demethylase in somatic tissues, and loss of ROS1 leads to hypermethylation of numerous discrete regions throughout the genome. *ROS1* is expressed broadly, including the endosperm. However, the role of ROS1 in endosperm, if any, has been largely unexplored. We used whole-genome enzymatic-methyl sequencing (EM-seq) of wild-type and *ros1* mutant endosperm to assess the contribution of ROS1 to endosperm DNA methylation patterning. *ROS1* is required to prevent hypermethylation at several hundred sites in endosperm, including at the edges of highly methylated transposable elements. We found that a subset of ROS1 target loci have a reduced capacity for hypermethylation in the *ros1* mutant endosperm relative to *ros1* leaf. We hypothesized this might be due to biased effects of *ROS1* on either the maternally or paternally inherited genome. To test this hypothesis, we performed EM-seq on F1 hybrids of distinct Arabidopsis strains. Recent results suggest that *ros1* hypermethylation at these sites occurs primarily on paternally-inherited alleles. Given the established role of DME in maternal genome demethylation, future work will be focused on the potential for conflict between DME and ROS1 in the endosperm, and how this could inform our understanding of parental conflict and gene imprinting in the seed.

MECHANISTIC INSIGHTS INTO RNA-GUIDED DNA INSERTION OF TYPE V CRISPR-ASSOCIATED TRANSPOSONS

Irma Querques[1,2], Michael Schmitz[2], Seraina Oberli[2], Christelle Chanez[2], Martin Jinek[2]

[1]Max Perutz Labs, University of Vienna, Department of Structural and Computational Biology, Vienna, Austria, [2]University of Zurich, Department of Biochemistry, Zurich, Switzerland

Several Tn7-like transposons co-opted CRISPR-Cas machineries to direct transposon DNA insertion into target sites specified by CRISPR RNAs (crRNAs). Due to their RNA-guided DNA integration activity, CRISPR-associated transposons (CASTs) hold great promise as programmable, site-specific gene insertion tools. We employed an integrative structure-function approach to investigate type V CASTs that rely on the pseudonuclease Cas12k, the transposase TnsB, the zinc-finger protein TniQ and the ATPase TnsC.
Focused on the CRISPR-Cas12k complex and the TnsC protein filament, our research explored guide RNA arrangement, interaction with target DNA, and RNA minimization strategies. We discovered that the TnsC filament assembly induces structural changes in bound DNA, enhancing our understanding of the molecular interplay between CRISPR and transposons.

In most recent work, our research delved deeper into the structural basis of RNA-mediated DNA insertion in type V CASTs and revealed an unprecedented interaction between the CRISPR machinery and the ribosomal protein S15. We uncovered a novel moonlighting role of S15 in prokaryotic biology beyond its traditional involvement in ribosome assembly. S15 was thus identified as an essential regulatory component of type V CASTs. This discovery's implications for genome editing were substantial, leading to the reconstitution of these systems in human cells.

Altogether, our studies uncover key mechanistic aspects of RNA-guided DNA transposition and offer prospects for repurposing CASTs as next-generation genome editing tools for targeted DNA insertion.

MOLECULAR BASIS FOR TRANSPOSASE ACTIVATION BY A DEDICATED AAA+ ATPASE.

Álvaro de la Gándara*[1], Mercedes Spínola-Amilibia*[1], Lidia Araújo-Bazán[1], Rafael Núñez-Ramírez[1], James M Berger[2], Ernesto Arias-Palomo[1]

[1]CSIC, Centro de Investigaciones Biológicas Margarita Salas, Madrid, Spain, [2]Johns Hopkins University School of Medicine, Department of Biophysics and Biophysical Chemistry, Baltimore, MD

Transposases drive chromosomal rearrangements and the dissemination of drug-resistance genes and toxins. Although some transposases act alone, numerous classes rely on dedicated AAA+ ATPase subunits that regulate site selectivity and catalytic function through poorly understood mechanisms. Here, we have characterized a widespread and streamlined family of transposons, IS21, which contains two genes: the transposase, IstA, and a essential AAA+ regulator, IstB. We have used solution and cryo-EM studies to determine that isolated IstA and IstB self-assemble with their cognate nucleic acids into inactive complexes to prevent futile ATPase cycles and deleterious chromosomal breaks. Moreover, a 3.6 Å-resolution structure of a ~1 MDa transpososome complex shows how the ATPase-regulator uses nucleotide-controlled assembly and DNA deformation to enable site selectivity, transposase recruitment/activation, and integration. Importantly, specific transposase/regulator interactions stimulate ATPase activity and trigger a large conformational change on the transposase that positions the catalytic site with the correct configuration to perform DNA strand transfer. These studies help explain how AAA+ ATPase regulators – which are used by classic transposition systems and CRISPR-associated elements – can remodel their substrate DNA and cognate transposases to promote function.

THE ASSEMBLY OF THE TN7 TARGETING COMPLEX BY A REGULATED STEPWISE PROCESS

Yao Shen[1,2], Shreya Krishnan[1], Michael T Petassi[3,4], Joseph E Peters[3], Alba Guarné[1]

[1]McGill University, Biochemistry, Montreal, Canada, [2]Present address: University of Oxford, Division of Structural Biology, Oxford, United Kingdom, [3]Cornell University, Microbiology, Ithaca, NY, [4]Present address: Columbia University, Biophysics and Molecular Biology, New York, NY

Tn7 transposable elements are known for their highly regulated integration mechanisms and programmability. Tn7 targeting pathways rely on specific target-site selection proteins (TnsD/TniQ and TnsE) and the AAA+ adaptor TnsC to recruit and activate the transposase (TnsAB) at specific sites. Combining cryo-EM imaging with biochemistry and biophysics, we show how TnsD recruits ADP-bound dimers of TnsC and acts as a nucleotide-exchange factor to release one protomer with exchange to ATP. This explains how TnsC from the prototypical Tn7 element assembles a heptameric ring unidirectionally from the target site. The stepwise assembly process is conserved in TnsE-directed insertions. Our results highlight how functionally distinct protomers within the heptameric TnsC ring provide checkpoints for target immunity and give insights into how insertions occur in a specific orientation.

IDENTIFICATION OF A CORE ALU RNA DOMAIN REQUIRED FOR RETROTRANSPOSITION.

John B Moldovan[1], John Yin[1], John V Moran[1,2]

[1]University of Michigan Medical School, Department of Human Genetics, Ann Arbor, MI, [2]University of Michigan Medical School, Department of Internal Medicine, Ann Arbor, MI

Alu elements mobilize via an RNA intermediate by a replicative processed termed retrotransposition, which has resulted in the accumulation of >1.1 million Alu copies in human DNA. Alu retrotransposition events contribute to human genetic diversity and, on occasion, can disrupt gene function, leading to sporadic cases of disease such as neurofibromatosis 1 and certain cancers. Alu retrotransposition requires the Long INterspersed Element-1 (LINE-1) open reading frame 2-encoded protein (ORF2p). Current models propose Alu RNA forms a ribonucleoprotein particle (RNP) containing signal recognition particle proteins 9 and 14 (SRP9/14) that can localize to ribosomes. Ribosomal localization then allows the Alu RNA poly(A) tract to effectively compete with the LINE-1 RNA poly(A) tail for ORF2p binding. However, whether other Alu RNA sequences are required for retrotransposition requires elucidation.

Here, we used a HeLa cell-based assay to identify Alu RNA sequences that are critical for retrotransposition. Remarkably, we discovered that Alu transcripts expressed from an RNA polymerase II (pol II) promoter can efficiently undergo a single round of retrotransposition. Inverse PCR confirmed the resultant Alu retrotransposition events exhibit hallmarks of LINE-1 ORF2p-dependent retrotransposition (i.e., short, flanking target site duplications, a 3´ poly(A) tract, and integration into a consensus LINE-1 ORF2p endonuclease cleavage site). Importantly, expressing Alu from a pol II promoter allowed us to bypass the strict requirement of intact A/B box promoter elements necessary for Alu transcription by RNA pol III. Thus, we could examine whether mutating RNA sequences within RNA pol II derived Alu transcripts affect retrotransposition.

Deletion mutagenesis experiments demonstrated that the first 46 nucleotides (nt) of the 7SL RNA Alu domain is necessary and sufficient for robust retrotransposition when expressed from an RNA pol II promoter. Consistent with current models, the 46 nt Alu RNA domain associates with SRP9/14 in HeLa cell extracts and can efficiently drive a single round of Alu RNA retrotransposition in HeLa cells. We speculate this Alu “core” domain forms a discrete RNA structure that allows for SRP 9/14 binding and ribosomal association, thereby allowing the Alu poly(A) tract to compete with the LINE-1 poly(A) tail for ORF2p binding. In sum, our data demonstrate that a minimal Alu RNA “core” domain of 46 nt has been critical for generating nearly 11% of human DNA.

A SUBSET OF NUCLEAR RECEPTORS INTERACTS WITH LINE-1 ORF2p TO REGULATE RETROTRANSPOSON INTEGRATION

Aurelien Doucet[1], Ramona Galantonu[1], Javier Garcia-Pizarro[1], Clément Monot[1], Larisa Okorokova[1], Margo Montandon[1], Nadira Lagha[1], Aurore-Cécile Valfort[1], Serdar Kasakyan[1], Pierre-Olivier Vidalain[2], Gael Cristofari[1]

[1]University Cote d'Azur, Inserm, CNRS, IRCAN, Nice, France, [2]University of Lyon, ENS de Lyon, Inserm, CNRS, CIRI, Lyon, France

LINE-1 (L1) retrotransposons are major drivers of mammalian genome plasticity and evolution. They replicate by a copy-and-paste mechanism through an RNA intermediate and a reverse transcription step occurring directly at chromosomal target sites, a process termed target-primed reverse transcription. However, how this is achieved in a chromatin context and coordinated with cellular activities remains largely unknown.
To gain insights into the interplay between the L1 machinery and its cellular host, we performed yeast 2-hybrid screens using a collection of L1-derived fragments as baits. We identified the estrogen-related receptor alpha (ERRα), a transcription factor belonging to the nuclear receptor family, as a cellular partner of L1 ORF2p, the catalytic component of the L1 replicative complex. While siRNA-mediated ERRα depletion marginally affects L1 retrotransposition, overexpression of an ERRα dominant-negative form inhibits this process. Through fluorescent two-hybrid assays in human cultured cells, we observed that several other steroid receptors, related to ERRα, including ERRβ and ERRγ, as well as the estrogen and glucocorticoid receptors, also interact with ORF2p, highlighting a high level of redundancy. To test whether these transcription factors may be involved in tethering L1 integration to specific genomic locations, we analyzed the presence of their DNA binding motifs around de novo L1 insertions obtained experimentally in cultured cells. Notably, binding motifs for several nuclear receptors were found enriched upstream of L1 insertion sites. Consistent with a tethering role for these transcription factors, anchoring the ERRα domain interacting with ORF2p to a large artificial repeat array lacking L1 integration motifs specifically inhibits L1 retrotransposition, suggesting that the L1 machinery is sequestered in a chromosomal region refractory to L1 integration.
Based on these observations, we propose that ERRα and several other related nuclear receptors might represent tethering factors that promote L1 access to particular genomic region. Collectively, these results link hormonal signaling pathways with the regulation of a major endogenous mutagen in mammals.

This work was supported by the European Research Council (ERC-2010-StG 243312, Retrogenomics), Agence Nationale de la Recherche (ANR-11-LABX-0028; ANR-15-IDEX-0001; ANR-21-CE12-0001), Institut National du Cancer and Inserm (Avenir 2008), Fondation pour la Recherche Medicale (EQU202403018069), and CNRS (GDR 3546).

The role of LEDGF in transcription is exploited by HIV-1 to position integration

Rakesh Pathak[1], Caroline Esnault[2], Rajalingam Radhakrishnan[3], Parmit Singh[3,4], Hongen Zhang[2], Ryan Dale[2], Abhishek Anand[1], Gregory Bedwell[3,4], Alan Engelman[3,4], Ali Rabi[5], Sahand Hormoz[6], Priyanka Singh[1], Henry Levin[1]

[1]NIH, NICHD, Bethesda, MD, [2]NIH, BSPC, NICHD, Bethesda, MD, [3]Dana-Farber Cancer Institute, Dept of Cancer Immunology and Virology, Boston, MA, [4]Harvard Medical School, Dept of Medicine, Boston, MA, [5]Massachusetts General Hospital, Dept of Surgery, Boston, MA, [6]Dana-Farber Cancer Institute, Dept of Data Science, Boston, MA

HIV-1 integration occurs across actively transcribed genes due to the interaction of integrase with host chromatin factor LEDGF. Although LEDGF was originally isolated as a co-activator that stimulates promoter activity in purified systems, this role is inconsistent with LEDGF-mediated integration across gene bodies and with data indicating LEDGF is a histone chaperone that promotes transcriptional elongation. We found LEDGF is enriched in pronounced peaks that match the enrichments of H3K4me3 and RNA Pol II at transcription start sites (TSSs) of active promoters. Our genome-wide chromatin mapping revealed that MLL1 had a dominant role in recruiting LEDGF to promoters and the presence of LEDGF recruits RNA Pol II. Enrichment of LEDGF at TSSs correlates strongly with levels of integration across the transcribed sequences, indicating that LEDGF at TSSs contributed to integration across gene bodies. Although the N-terminal Pro-Trp-Trp-Pro (PWWP) domain of LEDGF interacts with nucleosomes containing H3K36me3, a modification thought to recruit LEDGF to chromatin, we found H3K36me3 does not contribute to gene specificity of integration. These data support a dual role model of LEDGF where it is tethered to promoters by MLL1 and recruits RNA Pol II. Subsequently, LEDGF travels across genes to effect HIV-1 integration. Our data also provides a mechanistic context for the contribution made by LEDGF to MLL1-based infant acute leukemia and acute myeloid leukemia in adults.

STRUCTURAL MECHANISM OF LINE1 TARGET-PRIMED REVERSE TRANSCRIPTION

George E Ghanim[1,2], Thi Hoang Duong Nguyen[1]

[1]MRC Laboratory of Molecular Biology, Structural Studies, Cambridge, United Kingdom, [2]Present address: Princeton University, Department of Molecular Biology, Princeton, NJ

LINE1 and Alu elements are the most abundant retrotransposable elements within the human genome. While most are inactive, a small subset can still mobilize and generate new insertions by target-primed reverse transcription (TPRT). This mobility can be a significant genetic mutagen. Decades of research have illuminated many aspects of TPRT, in which a target DNA is nicked and primes reverse transcription of the retrotransposon RNA: a process catalyzed by the LINE1 encoded protein, ORF2p. However, ORF2p has posed challenges to biochemical and structural studies, thus the molecular understanding of how ORF2 executes TPRT remains poorly understood. Here we captured key structures of the human ORF2p-TPRT complex at up to 2.3 Å resolution using cryo-electron microscopy. These structures provide the first molecular views of the TPRT complex, uncover an unexpected arrangement of the target DNA, and explain the distribution of target site duplication lengths observed *in vivo*. Our results, together with AlphaFold3 predictions, offer key mechanistic insights into LINE1 retrotransposition, including cis-preference, target DNA selection, and TPRT.

CHASING THE JUMPING GENES

Zhao (ZZ) Zhang

Duke University, Pharmacology and Cancer Biology, Durham, NC

Activation of transposons, which make up ~50% of the human genome, is often equated to genomic parasitization. Their products (such as RNAs, proteins, and cDNAs) are largely considered to be benign, if not detrimental, to the host; and their mobilization generates DNA damage and mutations, thus being detrimental. Assuming that transposon activation is a very rare event that must be maintained under strict regulation, extensive efforts have led to the identification of mechanisms that silence transposons. Suspecting that our current understanding of transposons is incomplete, my lab is dedicated to build tools to monitor transposon activity in animals. In my talk, I will present our recent efforts on detecting transposon activity in human cells and discuss related implication for understanding human disease.

REPURPOSED RETROTRANSPOSONS PLAY ESSENTIAL ROLES IN BRAIN FUNCTION

Jason Shepherd

University of Utah, Neurobiology, Salt Lake City, UT

We discovered a new and unexpected intercellular communication pathway regulated by Arc, a neuronal gene critical for synaptic plasticity, memory, and cognition in mammals. Arc contains a Gag capsid homology domain that has conserved secondary structure with retroviruses that was co-opted from a distinct family of Ty3/mdg4 LTR retrotransposons. We found that Arc protein can self-assemble into virus-like capsids that are released from neurons in extracellular vesicles (EVs), which can transfer RNA and protein cell-to-cell. A number of Gag-like genes expressed in various cell types in mammals and humans are predicted to have similar homology suggesting that this form of intercellular communication may be present in multiple cell types and function in many biological systems. The Drosophila Arc (dArc) homologs also assemble capsids from a Gag domain that are released in EVs but originated independently from a distinct lineage of Ty3/mdg4 retrotransposons. These studies suggest that a diverse set of Gag genes have co-opted biochemical properties of ancestral retroelements to perform new host functions. However, their physiological functions and roles in human diseases are poorly characterized. In recent work, we found that the Gag-like paraneoplastic Ma antigen (PNMA) genes encode proteins that are also capable of capsid formation. PNMA proteins are associated with cancer-induced paraneoplastic syndromes that present with an autoimmune response and neurological symptoms. We found that PNMA2, which has also been co-opted from a Ty3 retrotransposon, encodes a protein that is released from cells as non-enveloped virus-like capsids. Recombinant PNMA2 capsids injected into mice induce autoantibodies that lead to learning and memory deficits. These observations potentially explain why PNMA proteins are associated with paraneoplastic disorders. The precise function and signaling pathways mediated by these repurposed retrotransposons promises to reveal new biology that underlies complex brain processes such as memory and will provide insight into neurological diseases.

TRANSPOSASE-DERIVED GENE ACTS AS ARCHITECTURAL PROTEIN TO REGULATE NEURONAL TRANSCRIPTION

Eirene Markenscoff-Papadimitriou, Natasha Mariano, Cedric Feschotte

Cornell University, Molecular Biology and Genetics, Ithaca, NY

The diversity of cell types in the mammalian brain is astounding, and the developmental mechanisms that generate this diversity are largely undiscovered. Here we explore the neurodevelopmental functions of a transposase-derived gene called POGZ. POGZ contains a DNA binding transposase domain derived from the pogo transposon, and a chromatin-binding zinc finger domain. We report the genomic binding of POGZ in the developing mouse and human forebrain at euchromatic loci and gene regulatory elements (REs). Using POGZ knockout mice, we show that POGZ promotes the active chromatin state and transcription of neuronal genes. Interestingly, POGZ-dependent genes and enhancers are located in a few hotspots in the genome. Using Micro-C methods applied to POGZ KO brain tissue, we observe a localized role of POGZ in mediating long-range enhancer-promoter interactions at POGZ hotspot loci. Our results reveal a novel neurodevelopmental function of a transposon-derived transcription factor in regulating genes and enhancers in mammalian developing neurons. POGZ is a high confidence autism risk gene and this research is funded by a grant from the Simons Foundation for Autism Research (SFARI).

REVERSE TRANSCRIPTASE-RELATED PROTEINS ARE INVOLVED IN ENVIRONMENTAL STRESS RESPONSE PATHWAYS

Irina A Yushenova[1], Fernando Rodriguez[1,2], Irina R Arkhipova[1]

[1]Marine Biological Laboratory, Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Woods Hole, MA, [2]Rochester Institute of Technology, Thomas H. Gosnell School of Life Sciences, Rochester, NY

Reverse transcriptase-related (*rvt*) genes are domesticated non-mobile reverse transcriptases (RTs), distinguished by their unusual presence in both eukaryotes and prokaryotes. They represent one of the rare cases of RT domain recruitment to perform cellular function in eukaryotic cells, joining telomerase RTs (TERT) along with spliceosomal Prp8, in which the RT moiety lost its polymerization activity. Phylogenetic analysis reveals that although *rvt* genes can be horizontally transferred, all bacterial *rvt* sequences are monophyletic and are clearly distinguishable from eukaryotic ones. Thus, *rvt* genes may have become domesticated before eukaryotes arose from a prokaryotic ancestor. Ecologically, *rvt* genes are found in free-living, often soil-dwelling or semi-aquatic organisms, and are prevalent in fungi. In most cases, *rvt* genes are single copy; preserve catalytic D'DD residues; are under purifying selection; and often contain introns. To discover the cellular function of RVT proteins, we applied various comparative genomic, transcriptomic, biochemical, structural, and functional approaches, as well as bioinformatic analyses of genomic and environmental datasets. Comparing results obtained for chosen bacterial, fungal and animal models, we conclude that RVT is involved in response to diverse environmental stresses via template-independent polymerization.

EVOLUTION OF IMMUNE SPLICING REGULATION MEDIATED BY TRANSPOSABLE ELEMENTS

Edward B Chuong

University of Colorado Boulder, BioFrontiers Institute, Boulder, CO

Intronic transposable elements (TEs) can harbor cryptic splice sites and become alternatively spliced as exons into host mRNA transcripts – a process termed “TE exonization.” However, while TE exonization events are abundant potentially a rich source of novel splicing events, most are expected to be weakly expressed or nonfunctional, and their biological significance remains largely unexplored. By surveying mammalian long-read transcriptomic datasets, we have identified hundreds of high-confidence TE exonization events with robust evidence of expression. Many are predicted to form alternative protein isoforms that are novel or poorly characterized, but may represent functionally important splicing events. Our functional studies provide evidence that many TE exonization events form truncated protein variants that affect multiple innate immune signaling pathways, including interferon signaling. These data support a broad role for TEs in facilitating the evolution of species-specific isoforms with immune regulatory functions.

TRANSPOSON-SUPPORTED REGENERATION IN PLANARIANS

Hae-Lim Lee[1], Axel Poulet[1], Sudheesh Allikka Parambil[1], Josien C van Wolfswinkel[1,2,3,4]

[1]Yale University, Dept of Molecular Cellular and Developmental Biology, New Haven, CT, [2]Yale University, Yale Center for RNA Science and Medicine, New Haven, CT, [3]Yale University, Yale Stem Cell Center, New Haven, CT, [4]Yale University, Yale Center for Research on Aging (Y-Age), New Haven, CT

Transposons are commonly viewed as selfish elements that are detrimental to their hosts. They are prime targets of specialized host defenses that constrain their expansion. Such defenses are particularly clear in long-lived stem cell systems such as in the highly regenerative planarian (flatworm) *Schmidtea mediterranea*. Our study of these defense mechanisms has identified an intricate interplay between transposon regulation and regenerative abilities in planarians. Planarian transposons are repressed by a combination of RNA- and chromatin-based strategies that are initiated in the stem cells. Defects in transposon silencing however do not directly affect the viability of the stem cells, but rather lead to tissue-specific deregulation of transposon transcripts, failed cell differentiation and fatal tissue dysfunction. This underlines the entanglement of transposons in developmental processes.

Remarkably, our studies also revealed an active transposon that confers a direct benefit to its host. We identified a Ty3-like transposon that entered the planarian genome over 100 million years ago, has actively expanded, and still retains its ability to mobilize. This transposon incorporated and optimized a copy of a host-derived anti-apoptotic protein. Transposon-based expression of this anti-apoptotic protein is upregulated upon stress and thereby improves stem cell survival. The presence of the transposon thus enhances the resilience of the stem cells, and this correlates with enhanced regenerative abilities of the planarians that host this transposon.

Our findings expand the range of interactions transposons have with their hosts to include mutualism. This provides a new perspective on the impact of transposable elements on eukaryote genomes, and on the adaptive host defenses against transposons.

COMPARTMENTALIZED RESPONSES TO ENDOGENOUS RETROVIRUSES GOVERN ADULT TISSUE REGENERATION

Ying Lyu, Soo Jin Kim, Ericka S Humphrey, Richa Nayak, Yejing Ge

MD Anderson Cancer Center, Department of Cancer Biology, Houston, TX

Constituting 40% of the genome, mammalian retrotransposons are largely domesticated for host benefit, with rare exceptions that remain selfishly active. The host has evolved a myriad of retrotransposon suppressors, many converging on H3K9me3 and DNA methylation. While adult soma mainly relies on the latter to repress retrotransposons, the former exhibits partly overlapping silencing functions whose specificity is less understood. During tissue regeneration, adult stem cells undergo major activation of lineage genes regulated by TET-mediated DNA demethylation. While TETs are known for their crucial role in DNA methylome remodeling during early development, their regulation of retrotransposons in adult tissues remains unclear.

We sought to tackle these questions in the murine skin, which harbors well-characterized, abundant, and genetically accessible adult stem cells. Hair follicle stem cells reside in an anatomically distinct niche, alternating between quiescence and activation in a synchronized fashion to fuel cyclic bouts of hair growth. Over repeated hair cycle, they eventually become exhausted with diminished regenerative capacity. The exhaustion-driven events of adult stem cells often remain elusive.

Here, we report a somatic pathway that couples adult stem cell activation with retrotransposon suppression to protect tissue regeneration. We observed dynamic expression of SETDB1 (H3K9me3 methyltransferase) closely coupled to stem cell activities in the murine skin. SETDB1 ablation led to reactivation of endogenous retroviruses (ERVs) and assembly of viral-like particles, resulting in hair loss and stem cell exhaustion. Importantly, the skin phenotype can be reversed by antiviral drugs, suggesting retroviral activity is functionally responsible for the skin pathology. Interestingly, we observed an unexpected, compartmentalized responses by different subsets of skin epithelial cells toward reactivated ERVs. On one hand, hair follicle progenitors mounted antiviral defense, and by deleting antiviral genes in SETDB1 deficient skin, we saw partially rescued phenotype. On the other, transient amplifying cells of the hair follicles did not induce antiviral genes despite ERVs being highly reactivated there. Instead, they were damaged through antiviral-independent mechanism: TET mediated DNA hydroxymethylation promoted ERV transcription and replication stress that subsequently drained the stem cell pool, a phenotype recapitulated by ablating cell fate transcription factors.

Together, we revealed compartmentalized epithelial responses against viral coding ERVs to protect adult tissue regeneration. Adult stem cells have adapted to live with selfish genes, and we speculate this is a long-sought molecular mechanism that drives stem cell exhaustion.

THE YIN YANG OF RETROTRANSPOSONS ACTIVITY IN THE SOMA: FROM HARMFUL TO RESILIENT RNAs IN AGING AND TISSUE REGENERATION

Valerio Orlando

KAUST, BESE, Thuwal, Saudi Arabia

Retrotransposons are mobile genetic elements that have colonized all eukaryotic genomes. While their role in evolution is widely accepted as a fundamental natural source of phenotypic variation, their activity in the soma is classically considered detrimental due to their potentially harmful effects on genome stability. However, recent studies have been unveiling previously unexpected regulatory and beneficial roles of retroelements reactivation at their RNA level. Indeed, although typically tightly negatively regulated, specific classes of retrotransposons, typically L1, and other repeat derived sequences have retained the ability to be reactivated especially in response to environmental cues, establishing an intriguing link between their expression and possibly major adaptive resilience functions. Conversely, retrotransposons and repeats mis-expression is well known to be associated with various acute stress responses, chronic inflammation and a hallmark of aging and associated diseases.
Hence, gaining insights in which signals regulate retrotransposons transcriptional homeostasis especially during life time, and the underlying physiological and possibly pathological mechanistic implications appear an important and timely endeavor.
In our lab we have successfully used either loss or gain of function approaches unveiling L1 RNA as target and tool to ameliorate accelerated aging (Della Valle et al, 2022) and bone homeostasis (Mangiavacchi et al, 2024), indicating retrotransposon RNA as pivotal regulators of resilience capacity.
Our work aims to extend the well-established DNA based evolutionary central role of retrotransposons as genetic drivers of phenotype variation to key epigenetic players regulating somatic resilience capacity through their RNA counterparts, impacting health and aging quality.

Ref.
Della Valle et al, Science Transl. Med. (2022) DOI: 10.1126/scitranslmed.abl6057
Mangiavacchi et al, EMBO J (2024) https://doi.org/10.1038/s44318-024-00143-z
Highlight by Martinez, Seluanov and Gorbunova (2024) https://doi.org/10.1038/s44318-024-00175-5

REGULATION OF ACTIVE LINE1 ELEMENTS EXTENDS LIFESPAN AND HEALTHSPAN IN MICE

John C Martinez[1], Matthew Simon[1], Francesco Morandini[1], Lucinda Fitzgibbons[1], Sung Jae Bae[1], Natasha Sieczkiewicz[1], Seyed Ali Biashad[1], Maxfield Kelsey[2], Michael Meadow[1], John Sedivy[2], Andrei Seluanov[1,3], Vera Gorbunova[1,3]

[1]University of Rochester, Biology, Rochester, NY, [2]Brown University, Molecular Biology, Cell Biology, and Biochemistry, Providence, RI, [3]University of Rochester, Medicine, Rochester, NY

Long interspersed nuclear element 1 (LINE1 or L1) comprises nearly 20% of the human genome, making it the most abundant transposable element (TE) in the genome. Aberrant L1 activity has been implicated in numerous diseases, including certain cancers, auto-immune disorders, neuroinflammation, and aging-associated inflammation, or 'inflammaging.' To combat this, cells have evolved mechanisms to transcriptionally repress L1 expression, primarily through heterochromatin packaging. As organisms age and heterochromatin becomes dysregulated, L1 expression is activated leading to DNA damage and L1 cDNA formation in the cytoplasm. These cDNAs are recognized by cytosolic sensors, such as cGAS, resulting in the activation of STING and subsequent type-I interferon (IFN) production. This has led to the proposal that appropriate regulation of L1s may lead to improved health outcomes, particularly in fending off unwarranted inflammation. However, the long-term effects of L1 downregulation have yet to be characterized. To explore this question, we generated novel L1 knockdown (L1KD) mice containing a cassette encoding shRNAs targeting the evolutionarily youngest L1 subfamilies in mice, MdA and MdTf. We show that these mice display significantly improved health outcomes with age, and, ultimately, extended lifespan driven by a strong anti-inflammatory transcriptomic signature. Importantly, these mice do not display obvious morphogenic, behavioral, or cognitive impairments compared to their wild-type counterparts. Additionally, these mice display extended reproductive health and longevity, characterized by healthy gonad architecture late into life and the ability to reproduce at advanced age. While the link between reproductive longevity and organismal longevity is still poorly understood, these results support the hypothesis that maintenance of gonad health translates to overall organismal health. Taken together, these results show that appropriate long-term regulation of active L1 mRNA is a promising avenue in alleviating 'inflammaging,' extending reproductive health, and reducing the effects of increased frailty with age.

MORE TO KRAB THAN MEETS THE EYE: HOW CHASSÉ-CROISÉS BETWEEN TRANSPOSABLE ELEMENTS AND POLYDACTYL PROTEINS GENERATED MECHANISTIC NOVELTY DURING VERTEBRATE EVOLUTION

Didier Trono

Ecole Polytechnique Fédérale de Lausanne, School of Life Sciences, Lausanne, Switzerland

KRAB-containing zinc finger proteins (KZFPs) are DNA-binding factors that tether epigenetic complexes to specific genomic loci. KZFP genes emerged some 400 million years ago in the tetrapod ancestral genome, and close to 400 human KZFPs constitute the largest family of transcriptional regulators encoded by our species. In their vast majority, these proteins bind to and control the activity of transposable element (TE)-embedded regulatory sequences. Reciprocally, more than 95% TE subfamilies are recognized by at least one and most often several KZFPs. Long neglected by genomic studies partly due to the difficulty of analyzing repetitive sequences partly because their lack of evolutionary conservation was wrongly interpreted as a sign of lesser biological importance, TEs and their KZFP controllers are increasingly recognized as essential drivers of the evolution of gene regulatory networks, playing key roles in multiple aspects of human biology including embryonic genome activation, imprinting, gastrulation, embryonic lineage determination, spermatogenesis, hematopoiesis, development of the heart, brain and pancreas, neuronal differentiation and function, adipogenesis, management of oxidative stress, inflammation and immune responses. Whereas 80% human KZFPs are canonical family members proven or predicted to act as transcriptional repressors through the KRAB-mediated recruitment of a KAP1/TRIM28-nucleated heterochromatin-inducing complex, the rest displays other functionalities linked to the presence of variant KRAB or additional domains such as SCAN or DUF3669. In my presentation, I will describe how during vertebrate evolution a series of chassé-croisés between poly-zinc finger protein genes and transposon sequences generated mechanistic novelty that underlies unprecedented functions for these non-canonical KZFPs.

EFFICIENTLY UNCOVERING THE INDIVIDUAL ROLES OF KZFPs AND DOMESTICATED TRANSPOSABLE ELEMENTS WITH A TRIM28-CENTRIC STRATEGY.

Juliette C Davis, Diana Voicu, Haskan Kaya, Santiago Morell, Michael Imbeault

University of Cambridge, Department of Genetics, Cambridge, United Kingdom

The family of KRAB zinc-finger proteins (KZFPs) is the largest group of DNA-binding factors in tetrapods and is rapidly evolving, with >350 protein-coding members in humans. While the individual function of most human KZFPs is poorly characterized, it is known that their primary role is to epigenetically silence transposable elements (TEs). KZFPs bind TEs in a sequence specific manner and recruit TRIM28 which interacts with other proteins to add H3K9me3 at the locus, inducing silencing-associated heterochromatin. It was discovered following a large-scale survey of their binding sites that most human KZFPs target evolutionary conserved transposable elements that have lost their transposition potential. This suggests additional selection pressures aside from their role in preventing transposition of mobile elements. For example, there is accumulating evidence that KZFPs participate in the domestication process of TEs and impact when and where they can affect gene regulation. As they have varied expression patterns, they can serve as epigenetic switches to control the accessibility of domesticated TEs in a cell-context specific way. For example, ZNF808 has recently been identified as an essential primate-specific regulator of pancreas development via the regulation of MER11 elements.

Here I present a new large-scale strategy to uncover links between KZFPs, TEs and gene regulatory networks. We leverage the shared ability of most KZFPs to recruit TRIM28, which is essential for their role as epigenetic silencers. Combining a TRIM28-centric approach with a high-definition catalogue of KZFP binding sites reveals which KZFPs are active in a particular cellular context. We have produced such genome-wide maps of TRIM28 binding in multiple cell types using an optimized ChIP-seq protocol to understand which TEs are targeted for silencing. Next, we have endogenously tagged TRIM28 with an inducible protein degradation tag to quantify the impact of silencing on gene expression. Proof of concept in HEK293T and HCT116 cells shows rapid degradation of TRIM28 leads to transcriptional changes from derepressed TEs associated with specific KZFPs and nearby genes. For example, we detected two LTR41s bound by TRIM28, associated with a known KZFP binding site, linked to the expression patterns of ACTL8, a normally testis-restricted gene. Further investigations in multiple cell types will allow for a global understanding of the roles of domesticated TEs regulated by KZFPs, and their contribution to genome regulation in human health and biology.

TRANSPOSABLE ELEMENTS IN HUMAN HEALTH AND DISEASE

Douglas F Nixon[1], Helena Reyes-Gopar[1,2,3], Jez L Marston[4], Luis P Iñiguez[4], Bhavya Singh[4,5], Matthew Greenig[6], Miguel de Mulder Rougvie[4], Elsa Lawrence[6], Tongyi Fei[4], Nicholas Liotta[1,4], Ethel Cesarman[7], Cedric Feschotte[8], Manvendra Singh[8,9], Enrique Hernández-Lemus[3], Rodrigo R Duarte[4,10], Timothy R Powell[4,10], Nicholas Dopkins[1], Matthew L Bendall[4]

[1]Feinstein Institutes for Medical Research, Northwell Health, Institute of Translational Research, Manhasset, NY, [2]Universidad Nacional Autónoma de México, Programa de Doctorado en Ciencias Biomédicas, CDMX, Mexico, [3]Instituto Nacional de Medicina Genómica, Departamento de Genómica Computacional, CDMX, Mexico, [4]Weill Cornell Medicine, Division of Infectious Diseases, New York City, NY, [5]Icahn School of Medicine at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York City, NY, [6]University of Cambridge, University of Cambridge, Cambridge, United Kingdom, [7]Weill Cornell Medicine, Department of Pathology, New York City, NY, [8]Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY, [9]Max-Planck Institute, Max-Planck InstituteGöttingen, Germany, [10]King's College London, Department of Social, Genetic & Developmental Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, London, United Kingdom

We have investigated the role of transposable elements or endogenous retroelements in human health and disease. In humans, thousands of TEs remain transcriptionally active in terminally differentiated cells, where they likely possess undetermined roles. We have developed a software package, Stellarscope, which performs locus-specific TE quantification at a single cell level. At the core of Stellarscope is an algorithm that integrates statistical modeling, graph theory, and probabilistic reassignment to analyze sequencing data and estimate the amount of RNA (number of UMIs) originating from each TE locus included in the annotation. Using Stellarscope, we built an atlas of TE expression in human peripheral blood mononuclear cells (PBMCs). We found that locus-specific TEs delineate cell types and define cell subsets not identified by standard mRNA expression profiles. To investigate the role of TE expression in germinal cell (GC) cell fate determination, we used Stellarscope to study germinal center B cells. By integrating TE and coding gene expression profiles, we identified novel B cell sub-states and trajectories in the transitions between canonical dark zone, light zone, and differentiated B cells. We identified a novel model of affinity maturation occurring within the GC based on TE expression. Finally, we have developed a retrotranscriptomic wide association study (rTWAS) pipeline to determine relationships between genetic risk and TE expression, identifying novel robust TE risk mechanisms for psychiatric and neurodegenerative diseases. In summary, we have developed novel bioinformatic tools to study TEs in human health and disease. Funding: NIH grants CA260691, UM1AI164559

ENDOGENOUS RETROVIRAL ELEMENTS AWAKENING IN HEMATOPOIETIC CELLS REVEALS NOVEL REGULATORY ROLES OF A YOUNG ELEMENT

Meijuan Chen[1,2,3], Brian Zhang[1,2,3], Kun Wu[1,2], Mikhail Trostnikov[1,2], Si Liu[1,2,3,4], Reuben Franklin[1,2,3], Shiyang He[1,2,3], Jernej Murn[1,2,3], Dustin Schones[5], Maria Ninova[1,2], Sihem Cheloufi[1,2,3]

[1]Department of Biochemistry, University of California, Riverside, CA, [2]Center for RNA Biology and Medicine, UCR, Riverside, CA, [3]Stem Cell Center, UCR, Riverside, CA, [4]Department of Microbiology & Plant Pathology, UCR, Riverside, CA, [5]Department of Diabetes Complications and Metabolism, City of Hope Beckman Research Institute, Duarte, CA

Endogenous Retroviruses (ERVs) are a class of heterogenous repetitive elements that are generally kept silenced throughout the life of an organism via epigenetic mechanisms. However, depending on when, where and which elements become activated, they can either present an important asset for gene regulation and evolution or pose a threat to the host. Identifying functional ERVs in somatic cells has been challenging in part due to their repetitive nature and the lack of cellular systems to study their regulation and function. Our work and others have previously identified the chromatin assembly factor-1 (CAF-1) complex as a critical regulator of ERVs during preimplantation development, in embryonic stem cells and during somatic cell reprogramming. More recently, we have demonstrated that CAF-1 depletion in hematopoietic stem and progenitor cells triggers their differentiation into a plastic mixed lineage state. Here we use this somatic cellular paradigm to assess the regulation of ERVs and the functional consequences of their activation. We found that CAF-1 controls chromatin accessibility at specific repetitive loci, including LINE1 and ERV elements. However, only ERV elements become transcriptionally active. Notably, activated ERV elements harbor cis-regulatory elements enriched in lineage-specific transcription factor motifs, suggesting they have become integrated into gene regulatory networks of hematopoietic cells. Furthermore, we find a remarkably strong and selective upregulation of a young ERV element. Characterizing these active copies led us to discover an ERV mediated host gene translational control and the production of virus-like particles with possible cellular roles. Our study suggests that manipulating chromatin accessibility in somatic cells provides a powerful paradigm to study the regulation and function of ERVs in various cellular contexts.

References: Manuscript in preparation, Franklin et al. Nat Commun 13, 2350 (2022) & Cheloufi et al. Nature 528, 218-224 (2015).

Funding: NIH R35 GM151004, the City of Hope/UCR biomedical research initiative & UC cancer research coordinating committee seed grants. MC & RF are supported by the California Institute for Regenerative Medicine Diversity fellowships.

MECHANISMS OF GENE SILENCING BY DNA METHYLATION

Steve Jacobsen[1,2]

[1]UCLA, MCDB, Los Angeles, CA, [2]HHMI, UCLA, Los Angeles, CA

DNA methylation is associated with transcriptional repression of eukaryotic genes and transposons. The mechanisms by which DNA methylation is established and maintained is well understood and involves a number of epigenetic feed forward loops involving histone methylation and non-coding RNAs. Our lab is studying the components and mechanisms by which DNA methylation is read and interpreted, and how it causes gene silencing or in some cases gene activation. We are also developing CRISPR based tools for the precise targeting of DNA methylation changes which can be used in basic plant research or in crop improvement projects.

THE REGULATORY POTENTIAL OF TRANSPOSABLE ELEMENTS IN MAIZE

Kerry L Bubb*[1], Morgan O Hamm*[1], Joseph K Min[1], Bryan Ramirez-Corona[1], Nicholas A Mueth[1], Jane Ranchalis[2], Mitchell R Vollger[2], Cole Trapnell[1,3,4], Josh T Cuperus[1], Christine Queitsch[1,3,4], Andrew B Stergachis[1,2,3,4]

[1]University of Washington, Department of Genome Sciences, Seattle, WA, [2]University of Washington, Division of Medical Genetics, Seattle, WA, [3]University of Washington, Molecular & Cellular Biology Program, Seattle, WA, [4]University of Washington, Brotman Baty Institute for Precision Medicine, Seattle, WA

The genomes of flowering plants consist largely of transposable elements (TEs), some of which can modulate gene regulation and function. However, the repetitive nature of TEs and difficulty of mapping individual TEs by short-read-sequencing has hindered our understanding of their regulatory potential. We demonstrate that long-read chromatin fiber sequencing (Fiber-seq) comprehensively identifies accessible chromatin regions (ACRs) and CpG methylation across the maize genome. We uncover stereotypical ACR patterns at young TEs that degenerate with evolutionary age, resulting in TE-enhancers preferentially marked by simultaneous hyper-CpG methylation and chromatin accessibility, a novel plant-specific epigenetic feature. We show that TE ACRs are co-opted as gene promoters and that ACR-containing TEs can facilitate gene amplification. Lastly, we uncover a pervasive epigenetic signature – hypo-5mCpG methylation and diffuse chromatin accessibility – directing TEs to specific loci, including the loci that sparked McClintock's discovery of TEs.

TRANSPOSON AMPLIFICATION IN ARABIDOPSIS EXPRESSING THE ONCOHISTONE H3.1K27M

Yannick Jacob

Yale University, Molecular, Cellular and Developmental Biology, New Haven, CT

The oncomutation lysine 27-to-methionine in histone H3 (H3K27M) is frequently identified in tumors of patients with diffuse midline glioma-H3K27 altered (DMG-H3K27a). H3K27M inhibits the deposition of the histone mark H3K27me3, which affects the maintenance of transcriptional programs and cell identity. Cells expressing H3K27M are also characterized by defects in genome integrity, but the mechanisms linking expression of the oncohistone to DNA damage remain mostly unknown. We now demonstrate that expression of H3.1K27M in the model plant Arabidopsis thaliana interferes with post-replicative chromatin maturation mediated by the H3.1K27 methyltransferases ATXR5 and ATXR6. As a result, H3.1 variants on nascent chromatin remain unmethylated at K27 (H3.1K27me0), leading to ectopic activity of TONSOKU (TSK), which induces DNA damage and transposon amplification. Elimination of TSK activity suppresses the genome stability defects associated with H3.1K27M expression, while inactivation of specific DNA repair pathways prevents survival of H3.1K27M-expressing plants. Overall, our results suggest that H3.1K27M disrupts the chromatin-based mechanisms regulating TSK/TONSL activity, which causes genomic instability in heterochromatin and may contribute to the etiology of DMG-H3K27a.

THE GENETIC BASIS OF NATURAL DNA METHYLATION VARIATION ON TRANSPOSABLE ELEMENTS IN *ARABIDOPSIS THALIANA*

Pierre Bourguet[1], Magnus Nordborg[1], Frédéric Berger[1], Eriko Sasaki[2]

[1]Gregor Mendel Institute of Molecular Plant Biology GmbH,, Vienna, Austria, [2]Kyushu University, Faculty of Science, Fukuoka, Japan

Natural variation in genome-wide DNA methylation profiles is widely observed in plant populations and is often associated with climate. Such 'epigenome' variation is hypothesized to contribute to local adaptation by modulating gene expression and transposable element (TE) activities. However, the regulation of the epigenome and its roles remain unclear. To address these questions, we investigated the genetic effects on CG methylation over TEs as an inherited epigenetic mark. Through GWAS approaches with natural populations of *Arabidopsis thaliana*, we discovered that CG methylation variation in pericentromeric regions is significantly affected by a genetic variant of *Cell Division Cycle Associated 7* (*CDCA7*), a major chromatin modifier in vertebrates. Notably, a major *CDCA7* natural allele appears old, suggesting that *A. thaliana* experienced a reduction in genome-wide CG methylation levels immediately after speciation from its ancestors. In this presentation, we will discuss the features of the genetic and molecular basis of CG methylation and its evolution.

CENTROPHILIC RETROTRANSPOSON INTEGRATION VIA CENH3 CHROMATIN IN ARABIDOPSIS

Tetsuji Kakutani

University of Tokyo, Biological Sciences, Tokyo, Japan

In organisms ranging from vertebrates to plants, major components of centromeres are rapidly-evolving repeat sequences, such as tandem repeats (TRs) and transposable elements (TEs), which harbor centromere-specific histone H3 (CENP-A, also called CENH3 in plants). Complete centromere structures recently determined in human and Arabidopsis suggest frequent integration and purging of retrotransposons within the TR regions of centromeres. Despite the high impact of “centrophilic” retrotransposons on the paradox of rapid centromere evolution, the mechanisms involved in centromere targeting remain poorly understood in any organism. I will show our recent results characterizing “centrophilic” de novo integration of Arabidopsis retrotransposons and discuss their evolutionary dynamics.

WHO AND WHERE? UNCOVERING ACTIVE TRANSPOSABLE ELEMENTS AND THEIR CONTEXTUAL MOBILIZATION PREFERENCES IN ARABIDOPSIS

Leandro Quadrana

Institute of Plant Science Paris-Saclay, University Paris-Saclay - CNRS, Gif-sur-Yvette, France

Transposable elements (TEs) are ubiquitous DNA sequences capable of self-propagating across genomes. To limit their mutagenic potential, TEs are typically controlled by epigenetic mechanisms, including DNA methylation in both plants and mammals. However, TEs can eventually evade their epigenetic silencing and transpose. Using population genomics data of *Arabidopsis thaliana*, we previously found that each Arabidopsis ecotype contains a different set of ~40 mobile TE families (Quadrana et al., 2016 *eLife*). Nonetheless, only a few Arabidopsis TEs have been observed to transpose experimentally so far (Tsukahara et al., 2009 *Nature*, Mirouze et al., 2009 *Nature*, Quadrana et al., 2021 *Nat. Commun*.), raising questions about the genetic and environmental factors that control their activity. Here, I will present unpublished data on the identification of the complete set of experimentally mobile TEs in Arabidopsis, their dependency on different epigenetic silencing pathways as well as environmental triggers, and the developmental specificities for heritable transposition.

TRANSCRIPTION OF A CENTROMERE-ENRICHED RETROELEMENT AND LOCAL RETENTION OF ITS RNA ARE SIGNIFICANT FEATURES OF THE CENP-A CHROMATIN LANDSCAPE

Bryce J Chabot[1], Ruiyi Sun[1], Asna Amjad[1], Savannah J Hoyt[1], Cecile Courret[2], Amanda M Larracuente[2], Rachel J O'Neill[1,3], Barbara G Mellone[1,3]

[1]University of Connecticut, Molecular and Cell Biology, Storrs, CT, [2]University of Rochester, Biology, Rochester, NY, [3]University of Connecticut, Institute for Systems Genomics, Storrs, CT

Centromeres depend on chromatin containing the conserved histone H3 variant CENP-A for function and inheritance, while the role of centromeric DNA repeats remains unclear. Retroelements are prevalent at centromeres across taxa and represent a potential mechanism for promoting transcription to aid in CENP-A incorporation or for generating RNA transcripts to maintain centromere integrity. Here, we probe into the transcription and RNA localization of the centromere-enriched retroelement *G2/Jockey-3* (hereafter referred to as *Jockey-3*) in *Drosophila melanogaster*, currently the only *in vivo* model with assembled centromeres. We find that *Jockey-3* is a major component of the centromeric transcriptome and produces RNAs that localize to centromeres in metaphase. Leveraging the polymorphism of *Jockey-3* and a *de novo* centromere system, we show that these RNAs remain associated with their cognate DNA sequences in *cis*, suggesting they are unlikely to perform a sequence-specific function at all centromeres. We show that *Jockey-3* transcription is positively correlated with the presence of CENP-A, and that recent *Jockey-3* transposition events have occurred preferentially at CENP-A-containing chromatin. We propose that *Jockey-3* contributes to the epigenetic maintenance of centromeres by promoting chromatin transcription, while inserting preferentially within these regions, selfishly ensuring its continued expression and transmission. Given the conservation of retroelements as centromere components through evolution, our findings have broad implications in understanding this association in other species.

HARNESSING R2 RETROTRANSPOSON MACHINERY FOR SITE-SPECIFIC SAFE-HARBOR TRANSGENE ADDITION TO THE HUMAN GENOME

Kathleen Collins

University of California, Berkeley, MCB, Berkeley, CA

Achieving the full promise of gene therapy will require a safe and generalizable approach for transgene addition to the human genome. To complement endogenous locus editing or disruption using CRISPR/Cas approaches, we engineered a system to install autonomously expressed gene cassettes at a multicopy safe-harbor locus. We adapted single-ORF R2 retroelement proteins to insert sequences of our choice to their native 28S rDNA target site by target-primed reverse transcription (TPRT). TPRT has two concerted steps: target-site nicking to generate the cDNA synthesis primer, and by cDNA synthesis using the protein-bound RNA template. Although R2 phylogenetic diversity is widespread across multicellular eukaryotes, mammals lost endogenous R2 yet retain the conserved target site. We deliver two synthetic RNAs to human cells: an mRNA encoding an avian R2 protein and a template RNA with payload sequence of choice flanked by compact 5' and 3' domains. After the RNP complex of protein and template RNA recognizes the target site, TPRT generates the first-strand cDNA. Second-strand nicking and synthesis complete stable transgene insertion by a method we term PRINT: Precise RNA-mediated INsertion of Transgenes. This technology escapes myriad challenges inherent in current methods for transgene delivery and is broadly deployable, dependent only on the production and delivery of RNA.

Funding statement: This work was supported by NIH DP1 HL156819.

Conflict of interest statement: KC is an inventor on patent applications filed by University of California, Berkeley related to the PRINT platform and has equity in Addition Therapeutics, Inc., which licensed the UC Berkeley technology.

STRUCTURAL RNA COMPONENTS SUPERVISE THE SEQUENTIAL DNA CLEAVAGE IN R2 RETROTRANSPOSON

Pujuan Deng*[1,2], Shun-Qing Tan*[1,2], Jia Wang[1,2], Jun-Jie Gogo Liu[1,2]

[1]Beijing Advanced Innovation Center for Structural Biology & Frontier Research Center for Biological Structure, School of Life Sciences, Tsinghua University, Beijing, China, [2]Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China

Retroelements are the widespread jumping elements considered as major drivers for genome evolution, which can also be repurposed as gene-editing tools. Here, we determine the cryo-EM structures of eukaryotic R2 retrotransposon with ribosomal DNA target and regulatory RNAs. Combined with biochemical and sequencing analysis, we reveal two essential DNA regions, Drr and Dcr, required for recognition and cleavage. The association of 3' regulatory RNA with R2 protein accelerates the first-strand cleavage, blocks the second-strand cleavage, and initiates the reverse transcription starting from the 3'-tail. Removing 3' regulatory RNA by reverse transcription allows the association of 5' regulatory RNA and initiates the second-strand cleavage. Taken together, our work explains the DNA recognition and RNA supervised sequential retrotransposition mechanisms by R2 machinery, providing insights into the retrotransposon and application reprogramming.

STRUCTURES OF THE *SLEEPING BEAUTY* TRANSPOSOSOME DURING EXCISION AND INTEGRATION

Vladimir Arinkin, Mark Dowling, Orsolya Barabas

University of Geneva, Department of Molecular and Cellular Biology, Geneva, Switzerland

DNA transposons of the *Tc1/mariner* superfamily have successfully spread across a wide range of eukaryotic genomes. They rely on a “cut-and-paste” transposition mechanism, driven by a transposase with a DD(E/D) catalytic domain and two DNA binding domains for recognition of Terminal Inverted Repeats (TIR) at the transposon ends.
One family member, *Sleeping Beauty* (SB) is widely used in genetic engineering to insert genetic material into eukaryotic genomes. Despite its broad use, the structure of the SB transposition assemblies (transpososome) has remained elusive, limiting our understanding of its mechanism and rational advancements of genetic applications.
Here, we present two high-resolution structures of the SB transpososome that represent key steps of its movement. The Paired-End Complex (PEC, 2.6 Å resolution) reveals the SB transposase binding two transposon ends after excision, and the Strand Transfer Complex (STC, 2.9 Å resolution) represents transposon integration into the target DNA.
Both structures reveal a dimer of the SB transposase, with monomers binding conserved outer repeats from the TIR. DNA binding and sequence recognition is achieved by two helix-turn-helix DNA binding domains, PAI and RED, which insert into the major grove with the connecting linker threading along the minor grove.
The DDE catalytic domain (CAT) overall resembles a canonical RNase-H fold, but deviates by the insertion of a "clamp loop" between the first and second catalytic residues. This clamp loop, the linker connecting CAT and RED, and the PAI domain together form an extensive dimer interface, which promote intricate subunit cooperation within the transpososome and ensure proper DNA positioning at the catalytic site. In STC, the target DNA interacts with the CAT and adopts a sharply bent conformation.
Comparing the SB transpososome structure to Mos1 from the *Mariner* family demonstrates marked differences in DNA recognition, dimer interface, clamp loop, and the C-terminal protein region. Strikingly, in contrast to Mos1, the SB clamp loop participates in sequence-specific TIR recognition, providing another level of fidelity check. The structure further reveals a new *Tc1*-specific role of the transposase C-terminus in recognizing the transposon boundary.
These differences underscore the evolutionary plasticity and adaptation of transposases and provide insights into the unique features of SB transposition. Our structural data on the SB transpososome further provides a resource for understanding transposase hyperactivity, paving the way for designing advanced transposon variants for cutting-edge genome engineering applications.

SINGLE-STRAND DNA TRIPLEXES ARE COMPONENTS OF DIVERSE MOBILE GENETIC ELEMENTS

Christopher G King[1], Aya Narunsky[2], Seth E Lyon[1], Neil White[1,2], Ronald R Breaker[1,2,3]

[1]Yale University, Department of Molecular Biophysics and Biochemistry, New Haven, CT, [2]Yale University, Department of Molecular, Cellular and Developmental Biology, New Haven, CT, [3]Yale University, Howard Hughes Medical Institute, New Haven, CT

Unusual DNA structures which fold into configurations beyond the canonical B-form double helix are exceedingly rare in biology, yet when formed they have important consequences for genome stability, DNA repair, or gene expression. Numerous mobile genetic elements rely on single-strand DNA (ssDNA) structures to define transposon termini, replication origins, or sites of integration, all of which are recognized by associated transposases, replicases, or integrases. Recently we have discovered a vast array of naturally occurring ssDNA triplexes distributed throughout bacterial, archaeal, and viral genomes. In some clades of ssDNA viruses, these triplex structures appear to be conserved components of the genomic strand, perhaps functioning in replication or packaging. In bacteria these structures sometimes occur at many hundreds of loci within a genome, often in dyad form. Phylogenetic profiling reveals a previously uncharacterized bacterial endonuclease which is always flanked by ssDNA triplexes and whose presence is correlated with triplex copy number. We propose that this genomic cassette functions as a novel type of transposable element in which ssDNA triplexes define the ends of the element. Our analyses reveal broad evolutionary conservation of ssDNA triplexes in diverse mobile genetic elements and expand the range of known ssDNA structures used by biological systems.

A RETROTRANSPOSON IN CANCER : THE MARKER AND THE MUTATOR

Kathleen H Burns[1,2]

[1]Dana-Farber Cancer Institute, Department of Pathology, Boston, MA, [2]Harvard Medical School,, Boston, MA

The human genome is replete with repetitive DNA attributable to the activity of self-propagating genetic sequences. In humans, this landscape is dominated by retrotransposons that make new copies of themselves by first being transcribed to RNA intermediates and then reverse transcribed to cDNA that is ultimately integrated into the genome. Long interspersed element-1 (LINE-1)-encoded proteins are responsible for this process, which occurs both in the germline and in somatic tissues. Increased expression and animation of genomic LINE-1 sequences appear to be hallmarks of cancer, and can be responsible for driving mutations in tumorigenesis. LINE-1 sequences encode a 6-kilobase (kb), bicistronic RNA intermediate transcribed from an internal RNA polymerase II promoter. The first of its open reading frames (ORFs) encodes ORF1 protein (ORF1p), which forms an RNA-binding homotrimer. Here, I will review data that this protein can serve as a marker of malignancies both in tissue biopsies and in the peripheral blood of cancer patients. The second ORF encodes ORF2p, which encompasses endonuclease and reverse transcriptase domains. These act in a coordinated manner to generate de novo genomic LINE-1 insertions, mutating cancer genomes. Here, I will review published and preliminary data that L1-mediated mutagenesis generates frequent double stranded (ds)DNA breaks leading to chromosomal deletions and inciting structural chromosomal instability (CIN). These findings suggest that LINE-1 may commonly contribute to cancer development and that enhancing its DNA damaging effects may represent a therapeutic strategy.

LONG-READ SEQUENCING PINPOINTS THE GENOMIC SOURCES AND REGULATORY DETERMINANTS OF ENDOGENOUS RETROVIRUS EXPRESSION IN CANCER

Justin S Becker[1,2,3], Daniel Fridman[1,3], Kevin Dong[1,3], Gary Sun[1,3], Maheshwaran Natarajan[1,3], Bradley E Bernstein[1,3]

[1]Dana-Farber Cancer Institute, Department of Cancer Biology, Boston, MA, [2]Massachusetts General Hospital Cancer Center, Thoracic Oncology Group, Boston, MA, [3]Broad Institute of MIT and Harvard, Epigenomics Program, Cambridge, MA

Tumor-specific expression of endogenous retroviruses (ERVs) has emerged as an important mediator of anti-cancer immune responses. One such mechanism associated with tumor regression is the recognition of ERV-derived protein antigens by T cells or circulating antibodies. However, each human ERV subfamily includes tens to hundreds of genomic copies capable of encoding proteins; the high similarity of their sequences has obscured differences in their activity and might suggest similar regulation among family members. We have developed a novel technique, using long-read Nanopore sequencing, to selectively target ERV proviral sequences and measure their epigenetic state in a locus-specific manner. Our method produces high-coverage maps of the sequence, DNA methylation, and chromatin accessibility surrounding ERVs, allowing identification of active LTR promoters. In both cancer cell lines and untransformed cells from the same tissues, we have discovered unexpected diversity in the methylation states of individual ERV proviruses, even within the same subfamily. Strikingly, we find that a small subset (~1%) of ERV proviruses, distinguished by highly accessible LTR promoters, are the major sources of tumor-specific ERV expression. The selective activation of these loci is in part sequence-intrinsic, mediated by motifs for the binding of oncogenic and stress-related transcription factors, and can be detected in data from human tumors. We trace a validated ERV tumor antigen back to specific source loci, and we find that these loci encode peptide sequences that differ from both the previously reported source locus and the reference genome (corrected using our long-read sequencing data). Collectively, our findings reveal that a relatively small number of ERV insertions account for cancer-specific provirus expression. Accurate identification of ERV antigens requires determination of the specific derepressed loci and their private mutations, which will be critical to targeting these antigens for therapeutic purposes.

TRANSPOSABLE ELEMENT SILENCING IN DROSOPHILA: INSIGHTS FROM GERMLINE AND SOMATIC PATHWAYS

Abdou Akkouche[1], Azad Alizada[2], Aline Martin[1], Emma Kneuss[2], Susanne Bornelöv[2], Nolwenn Mouniée[1], Stéphanie Maupetit-Mehouas[1], Benjamin Czech Nicholson[2], Gregory J. Hannon[2], Emilie Brasset[1]

[1]Université Clermont Auvergne, Institute of Genetics, Reproduction et Development (iGReD), Clermont-Ferrand, France, [2]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, United Kingdom

Transposable elements (TEs) pose a significant threat to genome integrity, requiring robust silencing mechanisms across different tissues. The PIWI-interacting RNA (piRNA) pathway plays a crucial role in suppressing TE activity in animal germline through complex transcriptional and post-transcripitonnal regulation to ensure effective TE silencing. Most piRNAs derive from the expression of specific genomic regions called “piRNA clusters”, regions densely populated with full-length and truncated TEs. In Drosophila, certain TEs have evolved so that they can now be expressed exclusively in somatic follicular cells surrounding the germ line, with some even capable of infecting and transposing within germ cells. This evolutionary pressure has led to the emergence of a simplified piRNA pathway in somatic follicle cells, which is essential for efficiently silencing TE expression in these cells and protecting the germline from potential infection. In this presentation, we will explore the distinct yet complementary mechanisms regulating piRNA pathway activity in Drosophila germline and somatic cells, providing a comprehensive understanding of the molecular and epigenetic controls essential for effective TE silencing.

TRANSPOSABLE ELEMENT SMALL RNAs AND LONG RNAs in HUMAN BRAINS DURING AGING AND HUNTINGTON'S AND PARKINSON'S DISEASE.

Gargi Dayama[1], Shruti Gupta[1], Brianne Connizzo[2], Xiaoling Zhang[3], Adam Labadorf[4], Richard H Myers[4,5], Nelson C Lau[1,5]

[1]Boston University Chobanian and Avedisian School of Medicine, Biochemistry and Cell Biology, Boston, MA, [2]Boston University, Biomedical Engineering, Boston, MA, [3]Boston University Chobanian and Avedisian School of Medicine, Medicine Section of Biomedical Genetics, Boston, MA, [4]Boston University Chobanian and Avedisian School of Medicine, Neurology, Boston, MA, [5]Boston University Chobanian and Avedisian School of Medicine, Genome Science Institute, Boston, MA

Recent studies suggest a role for Transposable Elements (TEs) in animal aging and neurogenerative disorders, yet the extent of their involvement in humans is not fully understood. We examined the emerging hypothesis that TE small RNAs (smRNAs) and TE messenger RNAs (mRNAs) are dysregulated during human aging and diseases like Huntington's Disease (HD) and Parkinson's Disease (PD). In aging human brain tissues and aging mouse brains, we tracked TE smRNA signatures that reflect inflammatory and/or RNA interference (RNAi) responses. We first compared from BrainSpan Atlas young and aged human brain smRNAs and mRNAs and complemented it with young and aged normal mouse brain. We discovered significant expression change of TE smRNAs from LINE and SINE elements in human aging brains whereas mouse brains lacked any such marked change of TE RNAs. We further uncovered a natural elevation of sense TE smRNAs in human frontal cortex, and a significant negative correlation between the TE smRNAs and mRNAs in differential expression changes for human SINEs and LINEs, suggesting a regulatory effect. We also observed a more prominent TE smRNAs dysregulation in HD, while PD seems to have a stronger impact on TE mRNAs. This suggests a distinct TE mRNA and smRNA regulation pattern between the two neurological diseases, which might be linked to early onset of the HD disease compared to PD. In particular, we found four prominent TE targets L1ME, L1PA, AluY, and HERVK that are most frequently dysregulated in HD. Interestingly, the TRIM28 transcription factor previously linked to TE repression was down-regulated with age in human but not in mice, possibly explaining the lack of change in TE RNA expression in aging mice. Altogether, our study shows the modulation of TE smRNAs and mRNAs may serve as novel RNA biomarkers of brain region integrity and health.

HENMT1 RESTRICTS ENDOGENOUS RETROVIRUS ACTIVITY BY 2'-O METHYLATION OF 3'-tRNA FRAGMENTS

Josh I Steinberg[1,2], Wolfram Gruhn[3], Jack Desmarais[1], Daisy Rubio[1], Justin Kinney[1], Azim Surani[3], Andrea J Schorn[1]

[1]Cold Spring Harbor Laboratory, Gene Regulation and Inheritance, Cold Spring Harbor, NY, [2]Stony Brook University, Medical Scientist Training Program, Stony Brook, NY, [3]The Gurdon Institute, University of Cambridge,, Cambridge, United Kingdom

Endogenous retroviruses (ERVs) utilize host tRNA as a primer for reverse transcription and replication, a hallmark of long terminal repeat (LTR) retroelements. Their dependency on tRNA makes these elements vulnerable to targeting by small RNAs derived from the 3′-end of mature tRNAs (3′-tRFs), which are highly expressed during epigenetic reprogramming and potentially protect many tissues in eukaryotes. We found high levels of 3'-tRF in male and female mouse primordial germ cells correlate with elevated ERV burden at sex-specific time points. tRFs are expressed when heterochromatin levels decline, before the onset of piRNA production in males and in the absence of piRNAs in the female germline. 3'-tRFs inhibit ERV activity by blocking reverse transcription but also by post-transcriptional silencing. Due to the perfect sequence complementarity of 3'-tRFs to endogenous retroviral sequences, they have tens of thousands of targets in mammalian genomes. We conducted a massively parallel reporter assay to determine target site rules for 3'-tRFs. Moreover, we found 3'-tRFs are 2'-O methylated by the small RNA methyltransferase HENMT1 outside the germline and in germ cells, akin to piRNAs and small RNA classes in other organisms that have perfect sequence complementarity to their targets but evade target-directed degradation.

INSIGHTS INTO *DE NOVO* piRNA PRECURSOR FORMATION IN MAMMALS

Zuzana Loubalova[1], Parthena Konstantinidou[1], Franziska Ahrend[1], Aleksandr Friman[1,2], Astrid Haase[1]

[1]National Institutes of Health, National Institutes of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, [2]University of Maryland, Biophysics Graduate Program, College Park, MD

The piRNA pathway is a conserved small RNA silencing pathway using PIWI proteins loaded with small RNA molecules called piRNAs (PIWI-interacting RNAs). PiRNA pathways function in germ cells of animals to silence transposable genetic elements and protect genome integrity. Mutations in piRNA pathways universally result in impaired germ cell development and sterility. PiRNAs are generated from single-stranded RNA precursors that originate from large genomic intervals called piRNA clusters. The process by which new piRNA clusters emerge and adapt to novel transposable elements remains largely unclear. In this study, we developed a bioinformatic tool to systematically identify and characterize piRNA clusters across various organisms. Our findings suggest that the formation of piRNA clusters in mammals is linked to transcriptional readthrough. We propose a novel mechanism for the generation of new piRNA precursors in response to retroviral infections. Results from our study improve our understanding of how new piRNA clusters emerge to control genomic invaders and protect germ cell development.

THE ZEBRAFISH piRNA PATHWAY SILENCES DIVERSE TRANSPOSABLE ELEMENTS AND ZINC-FINGER GENES IN A *PIWI* DOSAGE-DEPENDENT MANNER

Satyam P Srivastav, Andrew G Clark, Cedric Feschotte

Cornell University, Molecular Biology and Genetics, Ithaca, NY

Animals neutralize the threat posed by transposable elements (TEs) with the PIWI-piRNA pathway. Current understanding of the vertebrate piRNA pathway is derived mainly from rodent models that possess few transpositionally active TE families. Zebrafish is an attractive model system for studying the piRNA pathway, as ~56% of its genome consists of an extensive diversity of TEs, including 225 families with signatures of recent transpositional activity. Here, we describe discrete functions of the PIWI-piRNA pathway in adult zebrafish gonads using a combination of genetics, RNA-seq, and small RNA-seq. First, using *piwi* mutants, we identify diverse targets of the piRNA pathway, which includes distinct TE superfamilies, endogenous viruses, and satellites. Second, we find that silencing of most targets of the piRNA pathway is piwi-dose dependent. In fish heterozygous for either *piwil1* or *piwil2* null mutation, ovary and testis exhibit derepression of >200 TE families and satellites, implying both *piwi* genes are haploinsufficient in zebrafish gonads, unlike other vertebrate models. In homozygous *piwil2* mutant ovary, similar quantities of TEs are derepressed, but surprisingly >100 TE families and >500 genes are repressed. On the other hand, in homozygous *piwi* mutant testis, TE derepression is greater than in heterozygous. We hypothesize that this ovary-specific repression effect is the result of ectopic activation of other silencing pathways upon loss of *piwil2*. Notably, we find that 'Fish N-terminal zinc finger' (FinZ) domain encoding genes, which are thought to transcriptionally repress TEs, are upregulated in in *piwi* mutant ovary. A group of >300 FinZ genes are direct targets of piRNAs in wildtype ovaries and FinZ gene copies are enriched in ovary-biased piRNA clusters. These results suggest that FinZ genes are normally repressed by piRNAs in the adult ovary and therefore activated upon *piwi* loss, which leads to ectopic silencing of many TEs and genes throughout the genome. Fewer than 50 FiNZ genes are targeted by piRNAs in testis, which also does not exhibit broad FinZ activation and ectopic repression of TEs and genes. Thus, piRNA-mediated repression of FinZ genes appears to be required for proper gene expression in ovary. In summary, our study provides the first in-depth view of the diverse targets of the piRNA pathway in zebrafish and uncovers an intriguing crosstalk between two genome defense systems.

TRANSPOSON SILENCING IN DUCKWEEDS, TINY PLANTS WITH BIG SURPRISES

Veronica Barragan-Borrero[1], Daniel Buendia-Avila[1,2], Laura Diezma-Navas[1], Rana Elias[1], Rodolphe Dombey[1,2], Arturo Mari-Ordonez[1], Arturo Ponce-Mane[1], Filipp Krasnovid[1,2]

[1]Gregor Mendel Institute of Molecular Plant Biology (GMI), Austrian Academy of Sciences, Vienna, Austria, [2]Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, Austria

In plants, as opposed to metazoan, small RNA pathways dedicated to maintaining epigenetic TE silencing are ubiquitously expressed in all tissues and not exclusively localized to the germline. In flowering plants, such pathway, known as RNA-directed DNA methylation (RdDM), is characterized by the production of 24-nt sRNAs from TE loci to guide the deposition of DNA methylation and the formation of heterochromatin. Duckweeds (Lemnaceae) are a unique monophyletic family of flowering plants characterized by evolutionary simplification of development and fast asexual growth through clonal propagation. In contrast to the model plant Arabidopsis, duckweeds do not constitutively express RdDM and have lost some pathway components, resulting in the loss of 24nt sRNAs and associated mCHH DNA methylation patterns. To understand TE regulation and genome stability during clonal reproduction, we investigated three species ranging from the most ancestral to the most evolutionary young. Our data shows that TE silencing is maintained in the absence of somatic RdDM. However, the epigenetic landscape varies with the degree of TE genome colonization and the age of TE insertions. In duckweeds with no recent TE activity, only few recent TE insertions are covered with DNA methylation and H3K9me2 silencing marks while most TE sequences are associated with H3K9me1 and lack of DNA methylation. On the other side, species with more recent TE activity display higher levels of DNA methylation and a combination of transcriptional and postranscriptional silencing to regulate TE. Hence, even in the absence of active RdDM, alternative sRNA pathways might contribute to the regulation of TE proliferation. Thus, duckweeds offer an interesting system to investigate unexplored mechanisms of TE silencing in plants and how TE-host dynamics shape epigenomes.

EVOLUTION OF THE RAG TRANSPOSASE/RECOMBINASE

Eliza C Martin[1], Yuhang Zhang[2], Chang Liu[3], Lorlane Le Targa[4], Louis Tsakou-Ngouafo[4], Yi-Hsien Su[5], Pierre Pontarotti[4], Andrei J Petrescu[6], David G Schatz[1,7]

[1]Yale University, Department of Immunobiology, New Haven, CT, [2]Shanghai Jiao Tong University, School of Life Sciences and Biotechnology, Shanghai, China, [3]Johns Hopkins University School of Medicine, Department of Biophysics and Biophysical Chemistry, Baltimore, MD, [4]Aix-Marseille Université, IRD, APHM, MEPHI, IHU Méditerranée Infection, Marseille, France, [5]Academia Sinica, Istitute of Cellular and Organismic Biology, Taipei, Taiwan, [6]Institute of Biochemistry of the Romanian Academy, Department of Bioinformatics and Structural Biochemistry, Bucharest, Romania, [7]Yale University, Department of Molecular Biophysics and Biochemistry, New Haven, CT

Jawed vertebrates have evolved a sophisticated adaptive immune system that relies on assembly of immunoglobulin and T-cell receptor genes by a reaction known as V(D)J recombination. V(D)J recombination is initiated by the RAG1/RAG2 endonuclease (RAG). RAG possesses transposase activity *in vitro* and *RAG-like (RAGL)* and *Transib* transposons, which resemble RAG in sequence and structure, have been identified widely in the genomes of eukaryotes. We are attempting to understand the evolutionary steps that led from *Transib* and *RAGL* to RAG and V(D)J recombination. We are particularly interested in understanding the evolutionary adaptations that resulted in RAG's exquisitely regulated DNA binding and cleavage activities and to its loss of transposition activity *in vivo*. We have identified new *Transib* and *RAGL* transposons in a wide variety of invertebrate genomes, suggesting novel intermediates in RAG evolutionary history. I will discuss our ongoing efforts to identify new *Transib* and *RAGL* elements, to characterize the function of newly discovered RAG-family transposases and their domains, and to understand the mechanism by which RAG performs transposition.

SYNCHRONOUS L1 RETROTRANSPOSITION EVENTS PROMOTE CHROMOSOMAL CROSSOVER EARLY IN HUMAN TUMORIGENESIS

Sonia Zumalave, Jose Tubio

[1]Mobile Genomes, Centre for Research in Molecular Medicine and Chronic Diseases (CIMUS), Santiago de Compostela, Spain

L1 retrotransposition is a significant source of genomic variation in human epithelial tumours, which can contribute to tumorigenesis. However, fundamental questions about the causes and consequences of L1 activity in cancer genomes remain unresolved, primarily due to the limitations of short-read sequencing technologies. Here, we employ multiplatform sequencing, with an emphasis on long reads, and developed novel computational methods to analyse a fine selection of 10 tumours exhibiting high rates of somatic retrotransposition, encompassing over 6000 events. We leveraged long read data to adopt an approach to trace somatic L1s back to the source element whence they derive, based on the identification of single nucleotide variants specific to each L1 locus. This method outperforms previous strategies, and reveals a novel panorama of L1 loci activity, allowing us to identify novel L1 source elements in cancer, missed in previous works, and properly characterize the patterns of activity of the L1 loci that drive cancer retrotransposition. Furthermore, we dissect the internal structure of somatic L1s, finding recurrent patterns in their configuration, revealing the genetic mechanisms that inactivate somatic L1s, thereby limiting the exponential escalation of retrotransposition in a tumour. Notably, we observe evidence for previously undescribed or poorly documented mechanisms, including one we coined "twin priming & switching". A hidden landscape of chromosomal aberrations emerges in the light of long reads, where reciprocal translocations mediated by L1 insertion represent frequent events. Resolution of L1 bridges' configuration elucidates the mechanisms of their formation, where typically two independent, but synchronous, somatic L1 insertions drive the reciprocal exchange between non-homologous chromosomes. We developed timing approaches to deal with insertions of retrotransposons, finding that all analysed tumours exhibit more than a hundred retrotransposition events before the first whole-genome doubling (WGD) event, including L1-mediated reciprocal translocations and other retrotransposon-mediated structural variation. This indicates that activation of somatic retrotransposition is not merely the consequence of the genomic chaos typically governing in later stages of tumour progression. Instead, it appears to be a mutational process triggered early during tumour development, which can be very active and a main cause of genome instability. Overall, these findings highlight L1 activity as a more significant contributor to tumour genome plasticity than previously recognized, extending its impact beyond simple insertional mutagenesis.

TNPB OF IS200/605 TRANSPOSABLE ELEMENTS: AN OVERLOOKED LINK IN TRANSPOSITION THAT HAS BEEN REPURPOSED IN CRISPR-CAS SYSTEMS

Virginijus Siksnys

Vilnius University, Institute of Biotechnology, Life Science Center, Vilnius, Lithuania

The IS200/IS605 family of insertion sequences are among most simple and ancient TEs that carry only the genes and flanking elements that are required for its transposition and regulation. Typically, they contain subterminal left end (LE) and right end (RE) elements that flank either *tnpA* and *tnpB* genes in various configurations, or isolated *tnpA* or *tnpB* genes. TnpA transposase is essential for recognition, excision, and reintegration of TE of IS200/IS605 family that often carry an accessory tnpB gene, which is dispensable for transposition. We have shown that TnpB is RNA-guided nuclease that cleaves the donor joint after transposon excision [1]. We proposed that TnpB-mediated DSB triggers homology-directed repair to reinstate the TE into its original site analogous to the group I intron homing promoted by intron encoded endonucleases. The demonstration of the RNA-guided dsDNA cleavage activity of TnpB provided the first direct experimental evidence that TnpB is a predecessor of Type V CRISPR-Cas family. The TnpB target site is composite and is comprised of the TAM sequence (analogous to PAM) recognized by TnpB and a variable sequence complimentary to the TnpB-bound RNA guide (analogous to crRNA). ISDra2 TnpB ternary complex structure [2] 1. revealed the mechanism of TAM recognition paving the way for the rational design of TnpB variants with distinct TAM preferences. We also explored the natural diversity of the extremely abundant TnpB nuclease family by characterizing the biochemical cleavage requirements of the large set of TnpB nucleases and found that they show distinct TAM preferences. Taken together these data expands our knowledge on transposition mechanisms of the IS200/IS605 family and provides the platform for engineering of novel tools for the targeted genome modification.

1. Karvelis T. et al., Nature. 2021 599:692-696. doi: 10.1038/s41586-021-04058-1.
2. Sasnauskas G. et al. Nature. 2023 616:384-389. doi: 10.1038/s41586-023-05826-x.

ANTAGONISTIC CONFLICT BETWEEN TRANSPOSON-ENCODED INTRONS AND GUIDE RNAs

Rimantė Žedaveinytė[1], Chance Meers[1], Hoang C Le[1], Edan E Mortman[2], Stephen Tang[1], George D Lampe[1], Sanjana R Pesari[1,3], Diego R Gelsinger[1], Tanner Wiegand[1], Younggi D Moon[1], Samuel H Sternberg[1]

[1]Columbia University, Biochemistry and Molecular Biophysics, New York, NY, [2]Columbia University, Genetics and Development, New York, NY, [3]University of California, San Diego, Biochemistry and Molecular Biophysics, San Diego, CA

TnpB nucleases represent the evolutionary precursors to CRISPR-Cas12 and are widespread in all domains of life. IS605-family TnpB homologs function as programmable RNA-guided homing endonucleases in bacteria, driving transposon maintenance through DSB-stimulated homologous recombination. Here we uncover molecular mechanisms of the transposition lifecycle of IS607-family elements — found in all three domains of life — that, remarkably, also encode catalytic group I introns. We identify specific features for a candidate 'IStron' from *Clostridium botulinum* that allow the element to carefully control the relative levels of spliced products versus functional guide RNAs. Our results suggest that IStron transcripts evolved a sensitive equilibrium to balance competing and mutually exclusive activities that promote selfish transposon spread while limiting adverse fitness costs on the host. Currently, we are performing long-term evolution experiments to probe TnpB-mediated consequences of transposon retention, while also exploring regulatory control of TnpB expression via a broad family of genetically linked transcription factors. Collectively, this work highlights molecular innovation in the multifunctional utility of transposon-encoded noncoding RNAs.

TE MOBILITY DIFFERS IN SERIAL ISOLATES OF *CRYPTOCOCCUS NEOFORMANS* FROM RECURRENT HUMAN INFECTIONS

Anna Mackey, Vesper Fraunfelter, Samantha Shaltz, John Perfect, Sue Jinks-Robertson, Asiya Gusa

Duke University, Molecular Genetics & Microbiology, Durham, NC

Cryptococcus are species of environmental fungi that cause life-threatening diseases primarily in populations with weakened immune systems. *Cryptococcus neoformans* causes a fatal inflammation and swelling of the brain known as cryptococcal meningitis (CM), which is responsible for up to 20% of AIDS-related deaths. We previously reported TE mobilization as a significant driver of mutations in *Cryptococcus deneoformans* (a closely-related species) during murine infection and in response to heat stress in vitro, contributing to an increased rate of antifungal drug resistance at 37°C - human body temperature (Gusa et al, 2020). Subsequently, we demonstrated that TE mutations are the predominant source of genome-wide sequence variation under conditions of heat stress in *C. deneoformans*, compared to base substitutions or small insertions and deletions (Gusa et al, 2023).

To investigate TE mobilization in *C. neoformans*, the major disease-causing species, we obtained serial clinical isolates from patients with recurrent CM taken months apart. In this study, we identified several characterized and uncharacterized TEs (DNA transposons and retrotransposons) as mobile in *C. neoformans* under conditions of heat stress and compared the rate of drug resistance at 30° vs 37°. We found temperature-dependent differences in TE mobility as well as differences in TE mobility between serial isolates from the same patient. In order to understand these differences, we are constructing telomere-to-telomere genome assemblies to directly compare the genomes of serial isolates and identify genetic and genomic changes. Studies are underway to define the mechanism of heat stress-induced TE mobility in *Cryptococcus* and to determine whether there is evidence of TE movement during human infection.

LARGE SERINE INTEGRASES: HOW DO THEY KNOW WHICH WAY TO GO?

Heewhan Shin[1], Alexandria Holland[2], Abdulrazak Alsaleh[2], Alyssa D Retiz[1], Ying Z Pigli[1], Oluwateniola T Taiwo-Aiyerin[2], Tania Peña Reyes[1], Adebayo J Bello[2], Femi J Olorunniji[2], Phoebe A Rice[1]

[1]The University of Chicago, Biochem & Mol Bio, Chicago, IL, [2]Liverpool John Moores University, Pharmacy and Biomolecular Sciences, Liverpool, United Kingdom

Site-specific DNA recombinases catalyze DNA insertions, inversions, and deletions in an extremely tidy fashion, leaving no broken phosphodiester bonds. However, the mechanism by which they do so leaves them with an interesting thermodynamic problem: the net number of high-energy bonds in the product is the same as that in the substrate. How do these enzymes drive their reactions to near completion? Furthermore, how do they "decide" which pairs of sites to pair as substrates and in what relative orientation? I will describe our progress on answering these questions for the serine-family group of site-specific recombinases termed the large serine integrases. Large serine integrases are encoded by temperate bacteriophages, and have evolved to catalyze efficient insertion of large payloads. Furthermore, unlike tyrosine integrases, they do not require accessory DNA sites. Efficient insertion requires only a tetramer of integrase and two DNA sites: an ~40bp attB and an ~50bp attP site. Expression of the phage-encoded RDF (recombination directionality factor) protein, which binds the integrase itself, triggers the reverse (excision) reaction and inhibits integration. Our set of new cryoEM structures reveals how the RDF controls integrase conformation and controls reaction directionality.

We are also interested in expanding the toolkit of verified integrase – RDF pairs for use as synthetic biology tools. Surprisingly, although the RDFs encoded by a wide variety of phages are all predicted to bind to the same region of their cognate integrase proteins, they share no universally conserved sequence or predicted structural motifs, nor do their genes show consistent synteny with integrase genes. Finding RDFs by sequence alone has therefore been a bottleneck in the field. We found that AlphaFold2-multimer can be used to perform "virtual pulldowns" to identify putative RDFs. Wet lab testing of these predictions shows a high rate of true positives.

ACTIVITY OF THE MAMMALIAN DNA TRANSPOSON *PIGGYBAT* FROM *MYOTIS LUCIFUGUS* IS INHIBITED BY ITS OWN TRANSPOSON ENDS.

Alison B Hickman, Laurie Lannes, Christopher M Furman, Christina Hong, Lidiya Franklin, Fred Dyda

National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Laboratory of Molecular Biology, Bethesda, MD

piggyBat, identified in the genome of the little brown bat, is a member of the large *piggyBac* superfamily of DNA transposons and is to date the only known active DNA transposon found in mammals. *piggyBat* has a single open reading frame (ORF) encoding a 572 amino acid transposase that is flanked by 586 bp on its Left End (LE) and 324 bp on its Right End (RE). Although shorter ends are sufficient for activity, the sequences at the *piggyBat* termini do not appear to contain the same pattern of short repeated subterminal motifs as observed for *piggyBac*, its well-studied cousin from *Trichoplusia ni*, despite the 28.7% amino acid identity between the two transposases. To understand how the *piggyBat* transposase recognizes and acts on its transposon ends, we have used a combination of biochemical studies, in-cell transposition assays, and structural approaches. These have revealed that the low activity of *piggyBat* in human cells when compared to *piggyBac* is due to subterminal inhibitory DNA sequences, and that transposition activity can be dramatically improved by their removal. The cryo-electron microscopy structure of the *piggyBat* transposase pre-synaptic complex showed a mode of DNA binding distinct from that of piggyBac due to the use of topologically different C-terminal domains. Finally, the combination of structure-based re-engineering of the *piggyBat* transposase with LE and RE truncation resulted in a transposition system that gained about 100-fold increase in activity in HEK293T cells relative to wild-type *piggyBat*.

TARGETED GENOMIC INTEGRATION OF A LARGE DNA IN THE ABSENCE OF DNA DOUBLE STRAND BREAKS BY A PROGRAMMABLE MAMMALIAN TRANSPOSASE: INSERTION OF A THERAPEUTIC CFTR GENE INTO INTRON 1 OF THE CFTR GENE IN PRIMARY HUMAN CELLS

Nancy L Craig, Subhrangshu Guhathakurta, Jaclyn M Kubala, Sharmishtha Musalgaonkar, Daniel Pers, Rupesh Naidu, Esther Epum, Lillian Campos, Sarah Slauson, Joseph J Higgins, Joseph J Senn

SalioGen Therapeutics, Lexington, MA

The ability to insert large, therapeutically relevant genes into a chosen position in the human genome, for example, into intron 1 of a gene, such that the inserted gene is appropriately regulated by endogenous regulatory signals, has been long sought for gene therapy. Here we describe our Gene CodingTM technology platform for precise, targeted integration by transposition of large DNA segments into the human genome. Our enzyme, a bioengineered Myotis lucifugus transposase (bMLT), can efficiently integrate large (>12 kb) DNA cargos. Integration is mediated by a simple, single enzyme system in which a mammalian transposase is programmed to promote integration of a cargo DNA excised from a non-viral plasmid into a specific genomic TTAA site. Notably, our mammalian transposase, as opposed to others isolated from bacteria and insects, is well-adapted to function at the high growth temperatures of human cells.

A key feature of our Gene Coding technology is that integration occurs in the absence of potentially hazardous DNA double strand breaks in the genome, thereby avoiding the possibility of unrepaired DNA breaks and subsequent aberrant chromosome segregation, which can lead to chromothripsis and other highly mutagenic events. Insertion by transposition occurs via direct nucleophilic attack of the 3′OH transposon ends onto each target DNA strand.

With our Gene Coding technology, integration into a particular chosen TTAA genomic site is programmed by the fusion of our transposase to customized, sequence-specific DNA binding domains (ZFs/TALEs), which specifically bind to genomic sequences flanking the targeted TTAA. Notably, our transposase is engineered to dramatically reduce its intrinsic, non-sequence specific target DNA binding activity, which increases its dependence on the fused, sequence-specific DNA binding domains. This bioengineering step reduces insertion into positions other than the cognate one recognized by the fused DNA binding domains.

By NGS genomic sequencing, we have observed targeted ZFCFTR-bMLT directed integration at a particular targeted TTAA in Intron 1of CFTR in a G542X CF patient primary bronchial epithelial cell line with significantly reduced insertion elsewhere in the genome. Targeted integration in CFTR Intron 1 was also shown in up to 10% of transfected HEK293T cells by FACS.

COMPARATIVE GENOMICS, PHYLOGENETIC ANALYSIS, AND MODELING OF TRANSPOSABLE ELEMENT DYNAMICS ACROSS *CAENORHABDITIS* NEMATODE GENOMES

Victoria Eggers, Rohit Kapila, Gabriel Blanco, Sayran Saber, Jason Pienaar, Janna Fierst

Florida International University, Biological Sciences, Miami, FL

Transposable elements (TEs) have historically been referred to as genomic parasites, blamed for genome expansion, organismal sterility, and population extinction. However, many recent studies find TEs to be effectively neutral or even co-opted by the host genome to serve functions such as adaptive immunity or maintenance of genome architecture. Such conflicting findings are likely due to the diversity of TE dynamics. TE dynamics depend on transposition mechanism (Class I vs Class II and autonomous vs non-autonomous), family classification (sequence similarity), host defense systems, and host population dynamics (recombination rate, male frequency, and mating system). To investigate the interplay of these various factors, we first developed two deterministic models of autonomous and nonautonomous TE interaction, one in which the nonautonomous element parasitizes the autonomous one, and one in which the two compete for resources. We then compared real data to the simulations by annotating TEs in chromosome level assemblies of *Caenorhabditis* nematodes with EDTA and validating with EarlGrey. TEs were annotated as autonomous or nonautonomous based on the presence of required machinery for transposition. TE activity was assessed via sequence similarity (Kimura distance and coalescent models) and intersection with structural variants (TransposonUltimate). Phylogenetic analysis with Ornstein-Uhlenbeck and Brownian Motion models were investigated to decipher the intricate interplay of phylogenetic signal and nematode mating system against TE evolution. Finally, data from a *C.elegans* mutation accumulation experiment was analyzed for TE dynamics while also accounting for recombination rate and male frequency. Initial findings show class II mutator elements highly fragmented and mostly nonautonomous but maintained over phylogenetic time for both dioecious and androdioecious species, which supports a parasitic model of TE dynamics. Analysis from other transposon families and the mutation accumulation data is still in progress. This research elucidates the interaction of transposon dynamics and host population dynamics at multiple levels, furthering our fundamental understanding of genome evolution and evolutionary processes.

TOWARDS AN UNBIASED CHARACTERIZATION OF GENETIC DIVERSITY

Anna A Igolkina[1], Sebastian Vorbrugg[2], Fernando A Rabanal[2], Hai-Jun Liu[1], Haim Ashkenazy[2], Aleksandra E Kornienko[1], Detlef Weigel[2], Magnus Nordborg[1]

[1]Austrian Academy of Sciences, Gregor Mendel Institute, Vienna, Austria, [2]Max Planck Institute for Biology Tübingen, NA, Tübingen, Germany

Our view of genetic polymorphism is shaped by methods that provide a limited and reference-biased picture. Long-read sequencing technologies, which are starting to provide nearly complete genome sequences for population samples, should solve the problem—except that characterizing and making sense of non-SNP variation is difficult even with perfect sequence data. Here, we analyze 27 genomes of *Arabidopsis thaliana* in an attempt to address these issues, and illustrate what can be learned by analyzing whole-genome polymorphism data in an unbiased manner. Estimated genome sizes range from 135 to 155 Mb, with differences almost entirely due to centromeric and rDNA repeats. The completely assembled chromosome arms comprise roughly 120 Mb in all accessions, but are full of structural variants, many of which are caused by insertions of transposable elements (TEs) and subsequent partial deletions of such insertions. Even with only 27 accessions, a pan-genome coordinate system that includes the resulting variation ends up being 40% larger than the size of any one genome. Our analysis reveals an incompletely annotated mobile-ome: our ability to predict what is actually moving is poor, and we detect several novel TE families. In contrast to this, the genic portion, or ``gene-ome", is highly conserved. By annotating each genome using accession-specific transcriptome data, we find that 13% of all genes are segregating in our 27 accessions, but that most of these are transcriptionally silenced. Finally, we show that with short-read data we previously massively underestimated genetic variation of all kinds, including SNPs—mostly in regions where short reads could not be mapped reliably, but also where reads were mapped incorrectly. We demonstrate that SNP-calling errors can be biased by the choice of reference genome, and that RNA-seq and BS-seq results can be strongly affected by mapping reads to a reference genome rather than to the genome of the assayed individual. In conclusion, while whole-genome polymorphism data pose tremendous analytical challenges, they will ultimately revolutionize our understanding of genome evolution.

TRANSGENERATIONAL EPIGENETIC VARIATION AT TE SEQUENCES IN ARABIDOPSIS: MOLECULAR DETERMINANTS AND IMPACT IN NATURE

Pierre Baduel[1], Louna De Oliveira[1], Erwann Caillieux[1], Grégoire Bohl-Viallefond[1,2], Mounia El Messaoudi[1], Ciana Xu[1], Matteo Barois[1,4], Vipin Singh[1], Alexis Sarazin[1,5], Martine Boccara[1,7], Elodie Gilbault[3], Antoine de France[1], Leandro Quadrana[1,6], Olivier Loudet[3], Vincent Colot[1]

[1]ENS/PSL, IBENS, PARIS, France, [2]GMI, GMI, VIENNA, Austria, [3]Paris-Saclay University, IJPB, Versailles, France, [4]Lille University, EEP, Lille, France, [5]ETH, Biology, Zurich, Switzerland, [6]Paris Saclay University, IPS2, Gif-sur-Yvette, France, [7]MNHN, ISYEB, Paris, France

Transgenerational epigenetic inheritance (TEI) mediated by transposable element (TE) sequences is particularly prevalent in plants, presumably because of a limited reprogramming of DNA methylation between generations. However, the molecular determinants of this additional system of inheritance and the extent of its phenotypic impact in nature remain poorly characterized. Through a comprehensive analysis of severe loss of DNA methylation at TE sequences in an experimental population and >700 strains of *Arabidopsis thaliana*, we show that hundreds of loci across the genome sustain TEI in nature. Two main genomic features govern the prevalence of natural epiallelic variation at TE sequences: the number of cognate TE copies that can produce small RNAs to guide DNA methylation in *trans*, and gene proximity. Natural TE epialleles affect stress-responsive genes predominantly, suggesting that they are targets of selection. Thus, TE-mediated TEI constitutes an additional, genomically constrained source of heritable phenotypic variation with unique properties for plant adaptation.

JUNK DNA? MEGABASE-SCALE HETEROCHROMATIN REMOVAL ALLOWS FUNCTIONAL INTERROGATION OF *ARABIDOPSIS* PERICENTROMERIC REGIONS.

Valentin Joly, Máximo Kesselhaut, Sourav Roy, Yannick Jacob

Yale University, Dept. of Mol., Cell., & Dev. Biol., New Haven, CT

Recent telomere-to-telomere assemblies revealed that the *Arabidopsis* genome contains around 20% repetitive sequences. Besides centromeres, telomeres, and rDNA gene clusters, whose biological roles are well known, most of this repetitive material consists of transposable elements (TEs) concentrated in pericentromeric regions. A high density of constitutive heterochromatin, faithfully inherited at each mitotic cycle, ensures their robust transcriptional repression. For example, on chromosome 4, the centromere is flanked by two 800-kb regions essentially constituted of transposons repressed by H3K9me2 and DNA methylation. Each of those domains contains less than 10 expressed genes that are, for the most part, non-essential or uncharacterized. In several *Arabidopsis* accessions, the short arm of chromosome 4 also features a 450-kb heterochromatic knob generated by the well-known H4KS paracentric inversion. The large size of these regions, chromosomal localization, and remarkably low expression levels raise the question of their exact biological roles. Locally, does pericentromeric heterochromatin embed new, hidden functional loci? Globally, what is the structural role of pericentromeres in chromatin regulation and chromosome function?

To address this issue, we designed and optimized a new CRISPR/Cas9 chromosomal engineering strategy that allowed us to obtain large, heritable deletions of *Arabidopsis* heterochromatin with up to 10% success in the T1 generation. By combining a genome-wide bioinformatic annotation of heterochromatic guide RNAs with a high-efficiency Cas9 variant and a fast transformant selection method, we obtained a collection of homozygous deletions ranging from 300 kb to 800 kb in the pericentromeric and knob regions of chromosome 4. Surprisingly, most of those mutants were viable and fertile and showed no discernable phenotypes under normal growth conditions. Only the largest deletions showed either a retarded growth or a complete lethality that were rescued by restoring two isolated genes, respectively, the transcription factor *YY1* and the developmental gene *POPCORN*. By sequential genome editing, we were able to obtain a fully homozygous triple mutant featuring deletions of both pericentromeric domains and the knob, with a total size of 1.9 Mb. RNA-seq, BS-seq, and ChIP-seq analyses revealed a limited effect of this large-scale downsizing of constitutive heterochromatin on transcriptional and epigenetic states. This work's technical and biological findings shed new light on the structural plasticity of *Arabidopsis* chromosomes and the potential dispensability of heterochromatic regions.

PHASED T2T APE GENOMES SHOW DIFFERENTIAL TE INSERTION RATES

Mark Loftus[1], Ashley E Kirby[1], Parithi Balanchandran[2], Jessica M Storer[3], Weichen Zhou[4], The T2T Ape Consortium[5], Ryan E Mills[4], Evan E Eichler[5], Rachel J O'Neill[3], Christine R Beck[2], Miriam K Konkel[1]

[1]Clemson University, Genetics & Biochemistry, Center for Human Genetics, Clemson, SC, [2]University of Connecticut Health Sciences & Jackson Lab for Genomic Medicine, Genetics and Genome Sciences, Farmington, CT, [3]University of Connecticut, Molecular and Cell Biology, Storrs, CT, [4]University of Michigan Medical School, Human Genetics, Ann Arbor, MI, [5]University of Washington, Genome Sciences, Genome Sciences, WA

Retrotransposons are the only class of transposable elements (TEs) currently propagating in ape genomes. Here, we reconstruct the evolution of L1, *Alu*, and SVA in great apes since the radiation from gibbons roughly 24 million years ago. Following ancestral haploid genome reconstruction, structural variants unique to each species were retrieved from haplotype-resolved assemblies of five great ape species from the T2T ape consortium. We identified putative lineage-specific insertions and confirmed our findings with BLAT. Our results indicate considerable variation of L1 and *Alu* propagation rates between different great ape lineages. While *Alu* quiescence is further confirmed in orangutans, with only ~100 insertions per species over the last million years, the L1 mobilization rate is much higher in both Bornean and Sumatran orangutans (>3,000 L1 insertions each) compared to any other great ape. Furthermore, both species harbor at least 2½ times the number of L1s with intact open reading frames (>500 in each species) compared to all other great apes, indicating a high L1 mobilization rate. In contrast, common chimpanzee followed by bonobo harbor the lowest full-length L1 counts with intact open reading frames; though the number is substantially higher than previously reported, with 95 in common chimpanzee and 124 in bonobo. While most insertions do not cause harm, a few of them disrupt genes or regulatory regions. Lastly, we reconstruct the birth of one TE family, SVA, and its gradual rise in activity over millions of years in each great ape lineage. Using data from the Human Genome Structural Variation Consortium, we find that the previously reported SVA_F1 subfamily is establishing itself as a major driver of SVA expansion in the human population. Additionally, we determine that roughly 40% of SVA insertions harbor transductions (i.e., host sequence carried with the SVA to new genomic locations). Finally, we identify a small group of polymorphic SVA insertions that co-mobilize a whole processed pseudogene, and transductions with evidence for splicing within transcripts. Taken together, haplotype-resolved genome assemblies allow for the complete investigation of TE insertions including previously inaccessible genomic regions, allowing reconstruction of TE evolution across apes.

GENETICALLY VARIABLE REGULATION OF TRANSPOSABLE ELEMENTS IN DIVERSE MOUSE EMBRYONIC STEM CELLS

Arad Bustan[1], Haley Fortin[2,3], Catrina Spruce[3], Candice Byers[3], Christopher L Baker[1,2,3]

[1]University of Maine, Graduate School of Biomedical Sciences and Engineering, Orono, ME, [2]Tufts University, School of Graduate Biomedical Sciences, Boston, MA, [3]The Jackson Laboratory, Genetics, Bar Harbor, ME

Mouse embryonic stem cells (mESCs) capture the pluripotency of the pre-implantation blastocyst and express high transcriptome complexity including increased expression of transposable elements (TEs). Using short read RNA-sequencing to estimate TE expression, we found that mESCs derived from genetically diverse mice show variable levels of TE expression. These include MERVL elements associated with totipotency, and IAP and MusD elements that are still active in the murine genome. Using TE expression as a quantitative trait, we genetically map multiple loci regulating TE expression across different mouse backgrounds. These loci map to genomic regions with high sequence diversity among mouse strains, including large structural variants. For example, MERVL expression maps to a locus on Chr 10 near the *Duxf3* cluster of genes, known to play a role in converting mESCs to totipotent-like cells (2CLCs). Other loci contain clusters of genes encoding DNA-binding KRAB Zinc-finger proteins (KZFPs). KZFPs are expressed during early epigenetic erasure in the pluripotent epiblast, in part, to prevent mobility of TEs. Novel paralogs arising from structural variation, including duplications and inversions, allow evolutionary divergence of alleles that would be misaligned, miscounted, or ignored, using short-read sequencing in non-reference strains. To overcome these obstacles, and catalog the full breadth of transcript variation, we used PacBio long-read Isoseq platform to characterize diverse transcriptomes expressed in embryonic stem cells derived from C57BL/6J (B6) and DBA/2J (D2) mice, revealing a large cohort of uncatalogued transcripts. Full-length transcripts allowed us to distinguish coding differences between paralogous KZFPs arising from structural duplications in a non-reference strain. To validate these observations, we cloned two novel KZFPs from D2 strains and expressed them in B6 mESCs which lack these isoforms. We find that these novel KZFPs change chromatin accessibility at repeat regions, including a specific class of TEs. Together these observations highlight the complex transcriptional landscape among embryonic stem cells and highlight how rapidly evolving regions of the genome impact chromatin regulation, TE regulation, and alter gene expression in diverse genomes.

PHYLOGENETIC PROXIMITY RATHER THAN AQUATIC LIFESTYLE SHAPES PATTERNS OF HORIZONTAL TRANSFER OF TRANSPOSABLE ELEMENTS BETWEEN ANIMALS

Héloïse Muller[1,2], Rosina Savisaar[3], Jean Peccoud[4], Sylvain Charlat[5], Clément Gilbert[1]

[1]Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, Gif-sur-Yvette, France, [2]University at Albany, State University of New York, Department of Biological Sciences, Albany, NY, [3]Mondego Science, NA, NA, France, [4]UMR CNRS 7267 Ecologie et Biologie des Interactions, Université de Poitiers, Equipe Ecologie Evolution Symbiose, Poitiers, France, [5]LBBE - Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Villeurbanne, France

Transposable elements (TE) are transmitted vertically from parents to offspring through reproduction. TE can also undergo horizontal transfer and cross species barriers. Inferring and quantifying horizontal transfer events of TE among dozens of species is a computationally and methodologically challenging task because genomes contain many TE copies, scenarios alternative to horizontal transfers are difficult to reject, and assigning independent transfer events to specific branches of the host phylogeny is not trivial. For these reasons, formal tests on factors shaping global horizontal transfers of TE remain scarce. Here, we propose an approach to test the effect of structuring factors on the number of horizontal transfers of TE. Because several studies suggested that aquatic habitats may be more conducive to horizontal transfers than terrestrial habitats, we tested our new approach on this factor. For this, we sampled genomes from 126 terrestrial and 121 aquatic species distributed across 19 taxonomic groups, each of which underwent at least one habitat transitions along the metazoan phylogeny (including Chordata, Arthropoda, Annelida and Mollusca). This is the first study attempting to recover horizontal transfers of TE between animals at such a taxonomical scale. We inferred 6,043 independent events of horizontal transfer of TE across this dataset, which is three to six times higher than in previous large-scale studies. We found at least one event in 6,054 of the 23,419 pairs of species we investigated (25.8%). Regarding our factor of interest, even though we confirmed previous results in which the authors found an excess of horizontal transfers in teleost fish, aquatic species do not show more horizontal transfers than terrestrial ones across the 19 independent taxonomic groups. Thus, our results do not support the aquatic habitat as a major factor of horizontal transfers of TE. Finally, we used a Bayesian approach to statistically demonstrate that phylogenetic proximity is a major factor shaping horizontal transfers of TE in all four phyla. We also show that this effect is stronger for class I TE, and that it varies depending on phylogenetic groups.

TWO CENTURIES OF TRANSPOSABLE ELEMENT INVASIONS IN *D ROSOPHILA MELANOGASTER*

Almorò Scarpa[1,2], Riccardo Pianezza[1,2], Sarah Signor[3], Robert Kofler[2]

[1]Vienna Graduate School of Population Genetics, Population Genetics, Vienna, Austria, [2]University of Veterinary Medicine, Population Genetics, Vienna, Austria, [3]North Dakota State University, Biological Sciences, Fargo, ND

Largely based on the phenotypic effects generated by transposable elements (TEs), previous works revealed that three TEs invaded *Drosophila melanogaster* in the last century. Based on the genomes of historical specimens, more than 600 publicly available short-read data sets, and novel bioinformatics approaches, we substantially extend this invasion history. We show that a total of 11 TEs spread in *D. melanogaster* over the last 200 years. These 11 invasions increased the genome by ~1%, thereby accelerating the rate of genome evolution approximately 3800-fold compared to base substitutions. Leveraging data from over 1400 arthropod genomes, we suggest that the TE invasions were triggered by horizontal transfers, with *D. simulans* and species of the *D. willistoni* group acting as donors. We further show that TEs can penetrate populations extremely fast. For example, *Transib1* spread from three isolated epicenters in 2014 to all investigated populations within just two years. Our findings suggest that anthropogenic activities, facilitating habitat and population expansions of *D. melanogaster*, might have accelerated the rate of horizontal transposon transfer as well as the spread of the TEs into the worldwide population. Given the significant impact of TEs and the potential involvement of humans in their dispersal, our research has important implications for both evolution and ecology.

MICROSCOPIC BATTLE: DISCOVERING PHAGES AND BACTERIA IN THEIR NATURAL HABITAT

Frederique Le Roux[1], Karine Cahier[2], Damien Piel[2], David Goudenege[3], Yannick Labreuche[3], Pauline Daszkowski[2], Clarisse Labbé[2], Ruben Barcia Cruz [2], Bruno Petton[3], Marc Monod[4], Laurence Ma[4], Charles Bernard[5], Eduardo Rocha[5], Carine Diarra[1], Martin Lamarche[1], Manon Lang[1]

[1]University of Montreal, Microbiology, infectiology and immunology, Montreal, Canada, [2]Sorbonne Université, CNRS, UMR 8227, Integrative Biology of Marine Models, Roscoff, France, [3]Ifremer, PFOM, Plouzane, France, [4]Institut Pasteur, Biomic, Paris, France, [5]Institut Pasteur, Microbial Evolutionary Genomics, Paris, France

Facing the therapeutic impasse of antibiotics, farming systems, including aquaculture, should consider the extraordinary resource of phages, natural bacterial predators, for environmentally friendly practices. Sustainable and safe use of phages requires an understanding of their specificity and evolution. However, our knowledge of phage infection mechanisms is mainly based on model systems in laboratory conditions, which do not reflect the enormous diversity that exists in nature.

Using natural populations of marine bacteria (vibrios) infecting oysters, we have shown that most phages have a narrow host range. Their specificity depends first on their ability to bind to the host surface and second on their capacity to resist intracellular defense mechanisms. In the case of the oyster pathogen Vibrio crassostreae, we can track coevolutionary lineages in the marine environment. These are phage species capable of adsorbing specifically to clades nested within V. crassostreae. I will present preliminary results from a follow-up of these lineages in a new time-series sampling and genomic analyses that allow us to propose evolutionary scenarios.

Emerging resistance to phages in nature primarily results from horizontal gene transfer. We have discovered a family of phage satellites, named Phage-Inducible Chromosomal Minimalist Islands (PICMIs), which are widely distributed in the Vibrionaceae family. PICMIs are characterized by their reduced gene content, lack of genes for capsid remodeling, and the packaging of their DNA as a concatemer. These islands integrate into the bacterial host genome adjacent to the fis regulator and encode three core proteins essential for excision and replication. PICMIs rely on virulent phage particles to spread to other bacteria and protect their hosts from competitive phages without interfering with their helper phage. The discovery of PICMIs highlights the necessity of fully characterizing virulent phages and their host producers to prevent the spread of satellite-mediated resistance in phage therapy.

In conclusion, exploring natural populations of phages and bacteria is essential for developing fundamental knowledge, providing a solid basis for informed consideration of new therapeutic avenues in the fight against pathogens, while preserving the delicate balance of the microbial ecosystem.

DEFYING NORMS: SYSTEMATIC ANALYSIS OF piRNA CLUSTERS AND THEIR ROLE IN TRANSPOSABLE ELEMENT CONTROL REVEALS UNIQUE FEATURES IN DIFFERENT MAMMALS

Franziska Ahrend[1,2], Parthena Konstantinidou[1], Zuzana Loubalova[1], Alexandr Friman[1,3], Astrid D Haase[1]

[1]National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, [2]University of Regensburg, Biochemistry I, Regensburg, Germany, [3]University of Maryland, Institute for Physical Science and Technology, College Park, MD

Silencing of transposable elements is initiated by PIWI-piRNA pathways in germ cells. PIWI-interacting RNAs (piRNAs) originate from long single-stranded precursors encoded by genomic piRNA clusters. Our recently developed toolkit, the 'piRNA Cluster Builder' (piCB), identifies, prioritizes and characterizes piRNA clusters, laying the foundation for systematic comparison between species. Here, we systematically characterize piRNA clusters in mammals and identify unique features of human piRNA pathways. Through integration of bulk- and single-cell RNA-sequencing data, and piRNA sequencing data, we disentangled functionally distinct piRNA pathways. First, we performed a comprehensive gene expression analysis of the four human PIWI genes, PIWIL1, 2, 3, and 4 during germ cell development. We observed that PIWIL4 expression was not restricted to juvenile (pre-puberty) gonads but was also present in adult testes, alongside PIWIL1. This pattern diverges from that in other mammals, where these individual PIWI genes are not expressed simultaneously. Single-cell RNA-seq data revealed that PIWIL4 and PIWIL1 signals originate from different cell types: early stages of Spermatogonial Stem Cells (SSCs) and meiotic and post-meiotic spermatocytes, respectively. Finally, considering the specific length profiles of piRNAs associated with different PIWI proteins, we could attribute the activity of individual piRNA clusters to specific cell types. We observed three classes of piRNA clusters: The first class produced piRNAs specifically in SSCs in juvenile and adult testes, the second class produced piRNAs specifically in meiotic and post-meiotic spermatocytes and was unique to adult testes. Finally, a third class produced piRNAs in SSCs and meiotic spermatocytes in adult testes. Our current work aims to understand the function of these different piRNA clusters for transposon silencing and gene regulation.

TARGETING HERV-K BY ART DRUGS FOR THE TREATMENT OF PRIMARY EFFUSION LYMPHOMA

Anuj Ahuja[1], Vicenta T Alonso[1], Jesus D Mora[1], Enrique A Mesri[2], Ethel Cesarman[1]

[1]Weill Cornell Medicine, Department of Pathology and Laboratory Medicine, New York, NY, [2]University of Miami, Department of Microbiology and Immunology, Miami, FL

Human Endogenous Retroviruses (HERVs), remnants of ancient retroviral infections integrated into the human genome, have garnered attention for their potential roles in a variety of disorders and malignancies, though the precise nature of their contribution to many of them has yet to be elucidated. There are many types of HERV elements in the human genome, with multiple integrated copies of each, but some do encode functional proteins. An example is HERV-K, which has its own reverse transcriptase, protease, and integrase genes like other retroviruses. HERV-K elements have been shown to promote KSHV-induced oncogenesis. Antiretroviral therapy (ART), used to inhibit HIV lytic reactivation, has shown promise in slowing KS progression in AIDS patients. We hypothesized that ART drugs may also impact endogenous retroviral activity in cancer cells. To investigate this, we treated PEL cells with HIV protease inhibitors and observed a dose-dependent inhibition of cell proliferation. Notably, the IC50 of these drugs for the PEL cells proliferation was higher than the dose used to treat AIDS patients, but still clinically achievable without adverse effects. Our findings indicate that ART drugs induce apoptosis and autophagy in PEL cells while downregulating key regulators of cell cycle progression and survival, including cyclin D1, pAkt, and pERK. Additionally, we observed inhibition of HERV-K protease activity which is required for the maturation of HERV-K-Gag protein, and unexpected suppression of HERV-K-Env expression by these drugs. These insights suggest a potential therapeutic role of ART drugs in cancer management through targeted HERV-K inhibition. Further exploration of this mechanism may offer novel strategies for the treatment of KSHV-associated malignancies.

A POTENTIAL ROLE OF L1 AND ORF2P IN CANCER PREVENTION: PLANNING A CLINICAL TRIAL

Asya Akca Akyildiz[1], Tatiana Cajuso Pons[2], Nori Goto[3], Wenfeng An[4], Omer Yilmaz[3], Kathleen H Burns[2], Martin S Taylor[1]

[1]Brown University Warren Alpert Medical School, Department of Pathology, Providence, RI, [2]Harvard University, Department of Pathology, Boston, MA, [3]MIT, Department of Biology, Boston, MA, [4] South Dakota State University, Department of Pharmaceutical Sciences, Brookings, SD

Long interspersed element 1 (L1) is the only protein-coding transposon active in humans. One-third of the human genome was written by L1 through retrotransposition: a "copy-and-paste" mechanism catalyzed by its encoded ORF2 protein (ORF2p)'s endonuclease (EN) and reverse transcriptase (RT) activities. While L1 expression is repressed in normal tissues, L1 expression and activities have been implicated in the pathophysiology of cancer, autoimmunity, and aging, making ORF2p a potential therapeutic target. We and others have shown that RT inhibitors originally developed for HIV also inhibit ORF2p RT, and have shown promising results in model systems and clinical studies of colorectal cancer (CRC). Furthermore, HIV patients on RT inhibitors may have as much as a marked 40-50% reduced relative risk of colorectal, breast, and prostate cancers. Based on these findings and L1's known and emerging roles in DNA damage and innate immune signaling, we hypothesize that L1 and its ORF2p's RT activity contribute to carcinogenesis, which will be an attractive target for cancer chemoprevention.
To test this hypothesis, we focus on CRC carcinogenesis models, in which approximately 85% of cases proceed stepwise in the "chromosomal instability (CIN) pathway" with increasing CIN with progressive mutations that sequentially transform normal colorectal epithelium into adenomas, advanced adenomas, and invasive cancer. Engineered organoids with these mutations can recapitulate phenotypes of patient lesions in orthotopic mouse models, but are known to lack the CIN seen in corresponding patient lesions. We will present data showing patient lesions are L1 positive but that models are L1 negative. Therefore, L1 may be a missing factor in the biology of these models and in the carcinogenesis paradigm in general, where we hypothesize it causes CIN.

To test these hypotheses and work towards a potential clinical trial of RT inhibitors for cancer chemoprevention, we will express L1 in CRC organoids spanning the adenoma-carcinoma sequence and assess whether L1 can create CIN and drive progression both in vitro and in vivo in orthotopic mouse models of CRC. We assay the effects of EN and RT mutations and treat these models with NRTIs to ascertain the effects of L1 and ORF2p inhibition on tumor progression. We will also assess whether NRTIs reach the colorectal and can inhibit L1 in the correct tissue by using mouse retrotransposition reporter models driven in colorectal epithelium and in engineered colorectal retrotransposition reporter lines. If the hypothesis is supported, clinical trials will be conducted to prevent cancer in patients with genetic cancer syndromes such as FAP, with the aim of developing a chemoprevention strategy that can be then applied to the general population. NRTI combinations such as Truvada (FTC + Tenofovir) are very safe and effective in preventing HIV transmission in humans, providing an opportunity for repurposing and rapid adoption, if effective.

DESIGN AND APPLICATION OF AN L1 VECTOR WITH DUAL REPORTERS THAT MEASURE 5' TRUNCATION

Abhishek Anand[1], SePil Lee[1], Caroline Esnault[2], Hongen Zhang[2], Henry L Levin[1]

[1]Division of Cellular and Molecular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, [2]Bioinformatics and Scientific Programming Core, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

The human genome contains ~500,000 copies of the non-LTR retrotransposon LINE1, most of which are inactive. LINE1 ORF2p possesses endonuclease and reverse transcriptase activities which generate a single stranded nick at the site of insertion and at this position perform target primed reverse transcription. However, in 95% of LINE1 insertions 5' truncations occur that remove promoter and protein coding sequences. 5' truncation has played a significant role in shaping the genome as what is now a 3-gigabase genome would have otherwise been 8-gigabases. However, little is known about the mechanism of truncation. To identify factors responsible for 5' truncation we developed an assay that measures truncated and full-length insertions. We made a dual reporter system with GFP in the 3' UTR and mCherry in the 5' UTR of LINE1, where both genes are disrupted by an intron which is in the antisense orientation relative to LINE1 transcription. The reporters can only be expressed after splicing, reverse transcription, and integration of LINE1 into the genome. Full-length LINE1 insertions express red and green reporters, but if 5' truncated, only the green reporter is active. Comparing the ratios of cells with dual fluorescence to cells that only express green fluorescence using FACS reveals full-length insertions represent approximately 7% of all de novo integration. RNAseq analysis using this system showed that just 35% of the 3' GFP reporter was spliced compared to the 90% of the 5' mCherry reporter that was spliced. We also sequenced full-length and truncated insertions in HEK293T cells produced by the dual reporter to confirm the presence of integrated LINE1. Previous studies have identified two residues in LINE1 that appeared to alter the length of insertion. Using this dual reporter, we confirmed that mutations of these residues lowered the ratio of full-length insertions relative to 5' truncations by 2-fold. We have also made another version of the reporter with a constitutive promoter and integrated the entire construct in HEK293T cells using a Piggyback transposon delivery system. For genetic uniformity, we have made single cell derived clonal populations from these cells. In one clonal cell line, we screened approximately 95 candidate genes for contributions to 5' truncation. CRISPR mediated knockout of one candidate MOV10, showed a two-fold increase in full length insertions. MOV10 is an RNA helicase and is also known to be associated with decapping factors DCP1 and DCP2, suggesting a connection of 5' truncation with RNA decapping. We will be testing factors that interact with MOV10 and LINE1 RNA for contributions in 5' truncation.

TRANSPOSON CONTROL DURING ZEBRAFISH EYE REGENERATION

Krista M Angileri[1], Cristina Torres[1,2], Serah M Dureus[1,2], Nornubari A Bagia[1,2], Yu-Ching Liao[1], Alyssa F Moore[1], Cedric Feschotte[1]

[1]Cornell University, Molecular Biology and Genetics, Ithaca, NY, [2]Cornell University, College of Arts & Sciences, Biology Scholars Program, Ithaca, NY

Tissue regeneration is the process by which a damaged tissue is restored to its original structure and function. In vertebrates, regeneration success requires precise temporal control of inflammation and proliferation, which together create a cellular niche conducive to transposable element (TE) activity. TEs are self-replicating, mobile genetic elements found in high copy numbers within most eukaryotes, and when left unchecked, TE-encoded nucleic acid and/or protein products trigger an innate immune response, potentially disrupting host genome integrity. Interestingly, TEs are upregulated after injury in salamanders, sea cucumbers, and worms, and in salamanders, TE expression returned to homeostatic levels upon tissue repair, implicating active TE repression as a proponent for regeneration. The Piwi pathway, well-known for its robust ability to repress TE activity in animal gonads, including zebrafish, are required for planarian stem cell self-renewal and regeneration. ***We hypothesize that TE activation after injury constitutes a barrier to regeneration that must be overcome to achieve functional tissue restoration.***

The zebrafish retina is an ideal model for testing this hypothesis: 1) it can regenerate all major neural and non-neural cell types, 2) the process is robust and well documented, and 3) ~60% of the zebrafish genome is comprised of TEs, many of which are evolutionarily "young" and may retain transposition activity. Using publicly available RNA-seq datasets, we identified multiple TE subfamilies that are differentially expressed after ocular retinal pigmented epithelium (RPE) injury. Notably, eight TE subfamilies belonging to the BEL, MDG4, and ERV superfamilies were significantly upregulated at 2 and/or 4 days post injury (dpi), and by 7 dpi, their expression returned to near control levels. Moreover, 6 TE subfamilies were downregulated at 2 dpi suggesting some TE subfamilies are normally expressed during homeostasis. To determine whether components of the Piwi pathway are also upregulated during ocular regeneration, we assessed the in vivo expression patterns of *piwil1* and *piwil2* via *in situ* hybridization. Our preliminary data indicate that both *piwil1* and *piwil2* are upregulated in the RPE at 4 dpi compared to uninjured controls. Surprisingly, in the uninjured retina, *piwil2* loss-of-function mutants display aberrant retinal progenitor cell differentiation and/or proliferation, further implicating Piwi proteins as potential key regulators of somatic progenitor cell state in the zebrafish. Together, these data support our hypothesis that TEs become activated upon injury and must be controlled for successful regeneration.

MPP8, A COMPONENT OF HUSH COMPLEX, REGULATES DE-REPRESSION OF TRANSPOSABLE ELEMENTS DURING ANTIVIRAL RESPONSE

Shota Azuma[1,2], Yuka W Iwasaki[1]

[1]RIKEN, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan, [2]The University of Tokyo, Department of Biological Sciences, Graduate School of Science, Tokyo, Japan

Transposable elements (TEs) occupy a large portion of the genome and have often been regarded as “junk” because most of them lack the ability to transpose. However, recent research has revealed that TEs are involved in various physiological functions. In somatic cells, TE expression is repressed by factors such as TRIM28 (tripartite motif-containing protein 28) and the HUSH (human silencing hub) complex. It has been reported that many TEs were de-repressed in response to viral infection. During infection, TRIM28 loses its function through the loss of SUMO (small ubiquitin-like modifier) modification, and it triggers the de-repression of TEs. However, it is still unknown how other TE-repressive factors affect the de-repression of TEs during virus infection.
In this study, we investigated how the HUSH complex contributes to transposon de-repression during viral infection. Viruses encode a variety of proteins in their genomes, so different types of viruses induce different immune responses. To investigate the mechanism of virus type-independent de-repression, we first examined whether poly(I:C), a dsRNA analog, can induce TE de-repression. We found that transfection of poly(I:C) induced de-repression similar to that observed during viral infection. Next, we focused on the potential regulatory role of the HUSH complex in de-repression. The HUSH complex, consisting of TASOR, MPP8, and periphilin, is known to repress active TEs such as LINE-1. However, its contribution to TE de-repression during antiviral response has not been investigated. We examined how the HUSH complex is involved in poly(I:C)-induced TE de-repression. By reanalyzing public RNA-Seq data, we identified HUSH-regulated TEs and examined their expression patterns during poly(I:C)-induced antiviral responses. Although the HUSH complex is known as a regulator of TEs, we found that TEs regulated by MPP8, a HUSH component, tended to show a greater degree of de-repression. These findings suggest that MPP8 may possess an unknown mechanism that promotes TE de-repression during the infection response.

HIGH-THROUGHPUT INTERROGATION OF TnpB VARIANTS FOR TAM-DEPENDENT DNA CLEAVAGE

Marius Baltramonaitis, Virginijus Siksnys, Tautvydas Karvelis

Vilnius University, Life Sciences Center, Institute of Biotechnology, Department of Protein-DNA Interactions, Vilnius, Lithuania

RNA-guided nucleases from CRISPR-Cas bacterial defence systems are widely used in the genome editing field for their ability to efficiently introduce double-stranded DNA breaks in the genomic DNA of eukaryotic cells [1]. Despite their wide adoption, the large size of typical CRISPR-Cas9 and Cas12a nucleases (>1300 aa) significantly limits the delivery of their encoding genes to cells. Therefore, there is a demand for smaller nucleases. Recently, a new class of RNA-guided nucleases has been characterized [2]. The compact TnpB protein, found in the transposon of Deinococcus radiodurans, was proven to be an ancestor protein of the CRISPR-Cas12 nuclease family and a programmable RNA-guided endonuclease itself capable of efficiently cleaving genomic DNA [3]. However, the main drawback of ISDra2 TnpB is its relatively long TAM (analogous to PAM sequence for Cas proteins) sequence requirement (5'-TTGAT-3') for the DNA target sequence which reduces the spectrum of potential genomic targets. Determination of ISDra2 TnpB structure [4] revealed the molecular details of TAM recognition and made it possible to tackle this problem by using rational protein design. By adopting a high-throughput mutant library screening approach, we investigated TAM recognition by TnpB mutants, uncovering the molecular basis of TAM recognition and establishing principles for future engineering efforts. Overall, this study reveals the potential of a high-throughput framework to investigate TnpB variants for DNA cleavage.

1. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. Science 361, 866–869 (2018).
2. Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. Science 374, 57–65 (2021).
3. Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. Nature 599, 692–696 (2021).
4. Sasnauskas, G. et al. TnpB structure reveals minimal functional core of Cas12 nuclease family. Nature 616, 384–389 (2023).

SINGLE CELL RETROTRANSCRIPTOMICS WITH STELLARSCOPE: DEVELOPING A SINGLE CELL TRANSPOSABLE ELEMENT ATLAS OF HUMAN CELL IDENTITY

Helena Reyes-Gopar[1], Jez L Marston[2], Bhavya Singh[2], Matthew Greenig[3], Jonah Lin[4], Mario A Ostrowski[4], Kipchoge N Randall[2], Santiago Sandoval-Motta[5], Nicholas Dopkins[1], Elsa Lawrence[3], Morgan M O'Mara[2], Tongyi Fei[2], Rodrigo R Duarte[6,2], Timothy R Powell[6,2], Enrique Hernández-Lemus[5], Luis P Iñiguez[2], Douglas F Nixon[1], Matthew L Bendall[2]

[1]Feinstein Institutes for Medical Research, Institute of Translational Research, Manhasset, NY, [2]Weill Cornell Medicine, Division of Infectious Diseases, New York, NY, [3]University of Cambridge, University of Cambridge, Cambridge, United Kingdom, [4]University of Toronto, University of Toronto, Toronto, Canada, [5]Instituto Nacional de Medicina Genómica, Departamento de Genómica Computacional, Mexico City, Mexico, [6]Institute of Psychiatry, Psychology & Neuroscience, King's College London, Department of Social, Genetic & Developmental Psychiatry, London, United Kingdom

Single cell RNA sequencing (scRNA-seq) is revolutionizing the study of complex biological systems and diseases, yet most transcriptomic studies overlook the contribution of transposable element (TE) expression to the transcriptome. In both scRNA-seq and bulk tissue RNA sequencing (RNA-seq), quantification of TE expression is challenging due to high copy number and poorly characterized TE gene models. In recent years, several approaches have been developed for genome-wide TE expression profiling, including our approach for locus-specific quantification, Telescope. However, single cell data poses additional considerations that further complicate TE quantification. First, the number of reads per cell for scRNA-seq experiments is typically 1-3 orders of magnitude smaller than in RNA-seq, thus decreasing the amount of information that can be used to reassign multimappers. Second, droplet-based methods perform mRNA end tagging instead of full transcript sequencing, potentially reducing unique sequence content. In this work, we present Stellarscope, an analysis pipeline that reassigns multi-mapped reads to a single genomic locus by using an expectation-maximization algorithm. We show that confident reassignment can be achieved even when per-cell read counts are low by pooling information across cells while tracking cell-specific counts. We assess the performance of Stellarscope using matched RNA-seq and scRNA-seq datasets and examine TE expression patterns in several tissues including single-nucleus datasets from GTEx and multimodal data from peripheral blood mononuclear cells. We find that specific TEs delineate known cell types and can define cell subsets that were not identified by standard mRNA expression profiles. Stellarscope will enable new insights into the roles of specific TEs loci in transcriptomic heterogeneity, cellular mechanisms, and disease pathology. Funding: NIH.

A MULTI-OMIC CHARACTERIZATION OF HUMAN FIBROBLASTS OVEREXPRESSING LINE-1

Juan I Bravo[1], Christina D King[2], Joanna Bons[2], Samah Shah[2], Jacob Rose[2], Birgit Schilling[2], Judith Campisi[2], Bérénice A Benayoun[1,3,4,5]

[1]USC Leonard Davis School of Gerontology, Gerontology, Los Angeles, CA, [2]Buck Institute for Research on Aging, Gerontology, Novato, CA, [3]USC Dornsife College of Letters, Arts and Sciences, Molecular and Computational Biology Department, Los Angeles, CA, [4]USC Keck School of Medicine, Biochemistry and Molecular Medicine Department, Los Angeles, CA, [5]USC Norris Comprehensive Cancer Center, Epigenetics and Gene Regulation, Los Angeles, CA

Long interspersed element 1 (LINE-1; L1) are a family of transposons that occupy ~17% of the human genome. Though they are normally repressed in young somatic cells, numerous observational studies have noted an increase in L1 expression and/or L1 copy number with age and age-related conditions like cellular senescence. Importantly, L1 de-repression can engage the interferon antiviral response and promote the maturation of the senescence-associated secretory phenotype (SASP). Though much work on the mechanisms by which L1 can disrupt cellular homeostasis has been carried out, often in cancer or immortalized cell lines, comprehensive analyses of L1-driven cellular changes in non-cancerous, primary cells are limited. Here, we transiently overexpressed L1 in IMR-90 human lung fibroblasts and carried out mRNA-sequencing and mass spectrometry to characterize alterations in their transcriptome, cellular proteome, and secretome. Unexpectedly, we observed a suppression of the interferon pathway at the transcriptomic level and a significant downregulation of most of the detected secretome. This suppression of the interferon response was also observed following transient overexpression of L1 in WI-38 human lung fibroblasts. To further clarify the contributions of L1 to interferon signaling and contextualize our results within the broader literature, we profiled the effects of various L1 overexpression systems (low L1 expression, high L1 expression, high expression of the individual L1 open reading frames, and high codon-optimized L1 expression) in IMR-90 by mRNA-sequencing. Interestingly, overexpression of the individual L1 open reading frames, ORF1 or ORF2, consistently stimulated the interferon response, suggesting differential consequences of full-length and fragmented L1 overexpression in these cells. Our work adds to the limited studies on the cellular effects of L1 activation in non-cancerous, non-immortalized cells.

CHARACTERIZATION OF L1-RETROTRANSPOSON ENCODED PROTEIN ORF1p AND ITS RIBONUCLEOPROTEIN COMPLEX FORMATION

Raphael I Britt, John Sedivy, Gerwald Jogl

Brown University, MCB, Providence, RI

Long Interspersed Nuclear Element-1 (L1) is retrotransposon that comprises ~17 % of the human genome. The majority of L1 elements in the human genome are non-functional, however a small fraction of full-length and intact L1s retain their ability to retrotranspose. L1 encodes two proteins: ORF1p, which is a nucleic acid chaperone that coordinates the formation of the L1 ribonucleoprotein (RNP), and ORF2p, which has endonuclease and reverse transcriptase activities. To further investigate the dynamics of the L1 RNP, we expressed and purified a series of truncated HsORF1p proteins from E. coli. We then synthesized RNAs and combined them with purified HsORF1p. Using this workflow, we performed microscale thermophoresis (MST) binding experiments with a Nanotemper Monolith using fluorescently labeled RNA mixed with unlabeled protein as well as fluorescently labeled protein mixed with unlabeled RNA and determined that the HsORF1p-RNA binding affinity was ~1 nM. This agrees with experiments performed by Sandra Martin who used nitrocellulose filter binding assays with murine ORF1p and radio labeled RNA single-stranded DNA to determine binding affinity. Furthermore, we mixed ORF1p with RNAs of different lengths and demonstrated the RNPs can be isolated by analytical SEC. Lastly, we have begun visualizing the isolated RNPs by TEM to show the effects of RNA length on ORF1p RNP size and organization.

ADDING THE MISSING TILES TO THE PUZZLE: SEQUENCE RECONSTRUCTION OF KRAB-ZINC FINGER GENE CLUSTERS REVEALS MODES OF THEIR RAPID EVOLUTION IN MICE

Melania Bruno, Sharaf Maisha, Apratim Mitra, Ryan Dale, Todd Macfarlan

The Eunice Kennedy Shriver National Institute of Child Health and Human Development, The National Institutes of Health, Division of Developmental Biology, Bethesda, MD

KRAB-zinc finger protein (KZFP) genes constitute a rapidly evolving family of transcription factors, numbering in the hundreds in most mammalian genomes. Many evolutionary young and clade- or species-specific KZFPs bind to and repress endogenous retroviruses (ERVs). While it is known that ERVs establish themselves in genomes as remnants of exogenous retroviral infections and increase their copy number by retro-transposition, how new KZFP genes emerge to repress new ERVs is currently unknown.
In mice, where recent retroviral colonization of the germline has established several ERV families unique to mice, the KZFP gene family has similarly expanded, exhibiting parallel evolution across mouse strains. However, gaps in repeat-rich KZFP gene clusters in current genome assemblies have prevented a thorough reconstruction of young KZFP gene evolution.

We have generated de novo assemblies of the C57BL/6J and 129S1/SvImJ mouse strains using PacBio HiFi and Nanopore long read sequencing, allowing us to completely fill the gaps in all KZFP gene clusters sequence. We found that some KZFP gene clusters are notably larger and more heterogeneous between strains than previously thought, with one particular cluster on Chr4 measuring 5.4Mb and 6.9Mb in C57BL/6J and 129S1/SvImJ, respectively, relative to 2.5Mb described in the GRCm39/mm39 build. This cluster contains dozens of novel protein coding KZFP genes that display distinct DNA binding and ERV targeting patterns. Additionally, de novo assembly in *Mus spretus*, which diverged from *Mus musculus* only 1.7 million years ago, revealed a 2.4Mb cluster, indicating recent expansion in *Mus musculus*. Further comparison with *Rattus Norvegicus* and repeat annotation show lineage specific enrichment of mouse ERVs in this cluster, as well as in other KZFP gene clusters that either emerged or expanded in the mouse lineage, suggesting that new ERV acquisition might have increased the recombinogenic potential of these loci and facilitated the emergence of new KZFPs and divergence across strains. Our findings explain some strain-specific ERV regulation and exemplify how new ERV integrations can drive the evolution of new KZFPs that repress them.

DETERMINING THE GENETIC, ENVIRONMENTAL AND DEVELOPMENTAL BASIS OF HERITABLE TRANSPOSITION IN ARABIDOPSIS THALIANA

Mireia Bueno Merino, Basile Leduque, Pol Vendrell Mir, Leandro Quadrana

Institute of Plant Sciences Paris-Saclay, Gif-sur-Yvette, Gif-sur-Yvette, France

Transposable elements (TEs) are ubiquitous DNA sequences capable of self-propagating across genomes. To limit their mutagenic potential, TEs are typically controlled by epigenetic mechanisms, including DNA methylation in plants and mammals. Nonetheless, TEs eventually evade their epigenetic silencing and transpose. Using Arabidopsis thaliana, we recently reported that the mobilome of this species spans more than 150 TE families, and that the transposition rate in nature varies across ecotypes in association with genetic and environmental factors. Nonetheless, only a few Arabidopsis TEs have been observed to transpose experimentally so far, raising the question of what genetic and environmental factors control transposition. Here, we set out to measure the rate of heritable transposition using a comprehensive panel of genetically diverse Arabidopsis plants subjected to a range of environmental conditions. Overall, we detected significant heritable mobilization for at least 31 TE families, spanning both retrotransposons and DNA transposons. Importantly, we show that epigenetic perturbations are sufficient to trigger transposition bursts in a TE-specific manner, with the environment playing a secondary, modulatory role in TE mobilization. Unexpectedly, we found that some epigenetic mutants triggering extensive DNA hypomethylation, such as those affecting the chromomethylases CMT2 and CMT3, are dispensable for controlling transposition. To investigate the molecular mechanism underlying TE activity, we studied the developmental window where heritable transcription takes place, by collecting seeds derived from flowers from the same or different inflorescence, and from different stems. Overall, the segregation pattern of new insertions reveals that distinct TEs transpose at different reproductive stages, and that the environment shapes the developmental window in which transposition is occurring. Our work provides a first genetic and environmental (GxE) interaction map of transposition and sets the basis for uncovering the molecular mechanisms controlling the developmental specificity of TE reactivation.

TN7-LIKE TRANSPOSONS TARGETING CRISPR ARRAYS: UNVEILING A NOVEL FUNCTIONAL INTERACTION

Laura Chacon Machado, Joseph E Peters

Cornell University, Microbiology, Ithaca, NY

Transposons are diverse genetic elements capable of mobilizing within genomes, playing a significant role in generating diversity and facilitating evolution. Among transposons, Tn7 and related members of the Tn7-like family are renowned for their exceptional control over target site selection during transposition. Certain Tn7-like transposons have evolved to co-opt various CRISPR-Cas systems, enabling precise RNA-guided transposition. In our laboratory, the study of CRISPR-associated transposons has been a central focus, driving us to continuously scour genomic databases for new instances of Tn7-like genes adjacent to CRISPR-Cas genes.
In this study, we unveil a novel group of elements discovered in Cyanobacteria, wherein the transposons insert within a CRISPR array—an unprecedented arrangement suggesting an unexplored functional interaction between Tn7-like elements and CRISPR-Cas systems. Employing a bioinformatics approach, we characterized this novel group of transposable elements in terms of phylogeny, target site specificity and specialization. CRISPR systems represent an attractive target due to their presence in both genomes and plasmids, enabling transposons to fulfill the dual lifestyle characteristic of Tn7-like elements. Indeed, we identified elements targeting CRISPR arrays in plasmids and genomes. Intriguingly, phylogenetic analysis of the target selection protein, TniQ, revealed that the closest evolutionary relatives of these transposons target tRNA genes—common transposon targets due to their conserved nature. Notably, apparent loss of the C-terminal region of TniQ spanning approximately 50 amino acids corelates with the shift in targeting from tRNA genes to CRISPR arrays. Furthermore, we observed that these transposons possess a minimal gene repertoire, encompassing only the core genes essential for transposition. Interestingly, while they target the end of individual CRISPR repeats in the array, the actual insertion is directed approximately 45 base pairs upstream, within the spacer regions. This may allow the transposition event to avoid disrupting the host defense system, potential minimizing negative impacts on the CRISPR system. Despite our efforts, attempts to reconstitute these transposons in vivo in E. coli proved unsuccessful, indicating a potential requirement for additional factors encoded within the native host.
Overall, our findings contribute to the fundamental understanding of the intricate and diverse interplay between Tn7-like transposons and CRISPR-Cas systems in Cyanobacteria, the phylum housing the most types of CRISPR-guided transposons discovered to date.

ANCIENT ORIGINS OF *ENV* CARRYING RETROTRANSPOSONS

Shashank C Chary[1], Rippei Hayashi[1,2]

[1]John Curtin School of Medical Research, The Australian National University, Acton, Australia, [2]The Shine-Dalgarno Centre for RNA Innovation, The Australian National University, Acton, Australia

When a retrovirus integrates itself into the genome, its fate is to either be degraded into fragments, or become endogenized as a retrotransposon. Once they are endogenized, they must remain actively transposing, creating new copies in the genome, else it will eventually be lost. This requires the retroviral proteins, *gag* and *pol*, which most canonical retrotransposons consist of. However, a small class of Ty3/*gypsy* elements in drosophila, called errantiviruses, also carry a retroviral *env* gene, theorised to have been acquired from baculoviruses. These errantiviruses, when expressed, form virus-like particles in the somatic gonadal cells, which can infect the germline cells and create inheritable new genomic copies.
This has been assumed to be restricted to Dipteran species, however, we report that these *env*-containing *gypsy* elements are highly widespread across nearly every branch of metazoans, including ancient phyla such as Cnidaria, Ctenophora and Tunicata. Furthermore, we found that env-carrying endogenous retroviruses are similarly widespread, intact and multiplied in vertebrate genomes. Interestingly, structural features in pol are highly conserved between some classes of vertebrate elements, such as the recently identified ‘lokiviruses’, and errantiviruses found in ancient animals, suggesting their common origins. These observations indicate that *env*-carrying retroelements, *gypsy* errantiviruses and vertebrate endogenous retroviruses alike, have been thriving in metazoan genomes since before the split of bilaterian and non-bilaterian animals in the Precambrian era.

EXPLORING THE ROLE OF ENDOGENOUS RETROVIRUSES DURING HEMATOPOIESIS

Meijuan Chen[1,2,3], Brian Zhang[1,2,3], Kun Wu[1,2], Mikhail Trostnikov[1,2], Si Liu[1,2,4], Reuben Franklin[1,2,3], Shiyang He[1,2,3], Jernej Murn[1,2,3], Maria Ninova[1,2], Sihem Cheloufi[1,2,3]

[1]Department of Biochemistry, University of California, Riverside, CA, [2]Center for RNA Biology and Medicine, University of California, Riverside, CA, [3]Stem Cell Center, University of California, Riverside, CA, [4]Department of Microbiology & Plant Pathology, University of California, Riverside, CA

Endogenous retroviruses (ERVs) are molecular remnants of ancient retroviral infections, occupying a significant portion of mammalian genomes. ERVs are typically kept silent during developmental processes and normal homeostasis via RNA silencing and chromatin-based mechanisms. However, certain ERV elements can escape silencing, providing a source of genetic and epigenetic diversity to the host organism. The functions of ERVs when active during normal homeostasis remain poorly studied largely due the repetitive nature of these elements. Here we use a system of hematopoietic differentiation system to study the function of ERVs in cellular plasticity. Recent work from our lab uncovered that the suppression of CAF-1 in hematopoietic stem and progenitor cells (HSPCs) lead to their differentiation into a mixed lineage state by activating myeloid, erythroid and megakaryocytic fate genes. Given the reported role of CAF-1 in controlling ERVs during early embryonic development, we wondered whether it is also involved in regulating ERVs in this committed lineage differentiation system. Indeed, we detected selective chromatin opening and transcriptional activation of specific ERV subfamilies in CAF-1 depleted HSPCs. Intriguingly, analysis of the top active element revealed the expression of a number of copies encoding viral proteins. Furthermore, we detected chimeric ERV-host gene isoforms with possible roles in promoting cellular differentiation. Characterization of these isoforms suggests translational regulation of the host gene. Together our study provides novel insights into the biological roles of ERVs in hematopoietic cells and possible technological and therapeutic applications.

References

Manuscript in preparation, Franklin et al. Regulation of chromatin accessibility by the histone chaperone CAF-1 sustains lineage fidelity. Nat Commun 13, 2350 (2022) & Cheloufi et al. The histone chaperone CAF-1 safeguards somatic cell identity. Nature 528, 218-224 (2015).

Funding:
City of Hope/UCR biomedical research initiative & UC cancer research coordinating committee seed grants. SC is supported by NIH R35 GM151004. MC & RF are supported by the CIRM TRANSCEND fellowships.

ESCALATION OF GENOME DEFENSE CAPACITY ENABLES CONTROL OF AMPLICONIC SELFISH GENES

Peiwei Chen[1,3], Grace YC Lee[2], Alexei Aravin[1]

[1]California Institute of Technology, Division of Biology and Biological Engineering, Pasadena, CA, [2]University of California, Irvine, Department of Ecology and Evolutionary Biology, Irvine, CA, [3]Present address: Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY

From RNA interference to KRAB-ZFP guided chromatin silencing, diverse genome defense pathways silence selfish genetic elements to safeguard genome integrity. Despite their diversity, different defense pathways share a modular organization, where numerous specificity factors identify diverse targets and common effectors silence them. In the PIWI-interacting RNA (piRNA) pathway, which controls selfish elements in the metazoan germline, diverse target RNAs are first identified by complementary base pairing with piRNAs and then silenced by PIWI-clade nucleases via enzymatic cleavage. Such a binary architecture allows the defense systems to be readily adaptable, where new targets can be captured via the innovation of new specificity factors. Thus, our current understanding of genome defense against lineage-specific selfish genes has been largely limited to the evolution of specificity factors, while it remains poorly understood whether other types of innovations are required.

Here, we describe a new type of innovation, which escalates the defense capacity of the piRNA pathway to control a recently expanded selfish gene in *Drosophila melanogaster*. Through an in vivo RNAi screen for repressors of *Stellate*—a recently evolved and expanded selfish gene family—we discovered a novel defense factor, Trailblazer. Trailblazer is a transcription factor that promotes the expression of two PIWI-clade nucleases, Aub and AGO3, to match *Stellate* in abundance. Recent innovation in the DNA-binding domain of Trailblazer enabled it to drastically elevate Aub and AGO3 expression in the *D. melanogaster* lineage, thereby escalating the silencing capacity of the piRNA pathway to control expanded *Stellate* and safeguard fertility. As copy-number expansion is a recurrent feature of diverse selfish genes—including transposable elements, meiotic drivers, satellite DNA—across the tree of life, we envision that augmenting the defense capacity to quantitatively match selfish genes is likely a defense strategy repeatedly employed in evolution.

IDENTIFY CHEMICAL APPROACHES TO ENHANCE TRANSLATION OF LINE-1 ENCODED ORF2p FOR CANCER THERAPEUTICS

Wen-Chih Cheng[1], John D Heaps[1], Jaime H Cheah[2], Miguel A Franco [2], Jane K McIninch[2], Christian K Soule[2], Brittany Marion[2], Lisa N Miller[3], James Berstler[3], Kathleen H Burns[1]

[1]Dana-Farber Cancer Institute, Pathology, Boston, MA, [2]Broad Institute of MIT and Harvard, Center for the Development of Therapeutics, Cambridge, MA, [3]Broad Institute of MIT and Harvard,, Cambridge, MA

Transposable elements, known as “jumping genes”, are DNA sequences capable of moving within genomes. The Long INterspersed Element-1 (LINE-1; L1) is a type of retrotransposon that mobilizes via an RNA intermediate in a mechanism termed retrotransposition. L1 is the only active autonomous human transposon, encoding two proteins required for retrotransposition. Open reading frame 1 protein (ORF1p) is an RNA chaperone, and ORF2p encompasses endonuclease and reverse-transcriptase activities.

To limit retrotransposition, normal cells rely heavily on epigenetic silencing of L1 promoter via methylation. In contrast to normal cells, L1 promoters are commonly derepressed in p53-mutated epithelial cancers, including colorectal, pancreatic, lung, ovarian, breast, and prostate cancers. While hypomethylation in cancer leads to the expression of L1-encoded ORF1p, ORF2p expression is extremely limited in ORF1p(+) tumors. Because ORF2p is a potent inhibitor of cell growth and can sensitize cells to DNA-damaging chemotherapies, we hypothesize that cancer cells actively suppress ORF2p expression in ORF1p(+) tumors. Therefore, enhancing ORF2p expression in ORF1p(+) cells presents a therapeutic opportunity to target ORF1p(+) tumors specifically.

ORF2p is translated from a bicistronic transcript which requires ribosome re-initiation following ORF1p translation. We seek to enhance ORF2p translation via small molecule screening approaches. In collaboration with the Center for the Development of Therapeutics (CDoT) at Broad, a high throughput screening campaign was developed and executed using their highly-annotated Drug Repurposing library in a cell-based luciferase reporter system which monitors ORF2p production. Selected drug candidates were followed up in a secondary dose response screening that was recently completed. For small molecules validated in the secondary screening, we aim to determine their mechanism of actions in enhancing ORF2p production.

SUBFUNCTIONALIZATION OF TRANSPOSABLE ELEMENTS AS *CIS*-REGULATORY ELEMENTS ACROSS THE MOUSE IMMUNE SYSTEM

Jason D Chobirko, Elizabeth A Fogarty, Cedric Feschotte, Andrew Grimson

Cornell, Molecular Biology and Genetics, Ithaca, NY

The immune system is a rapidly evolving defense mechanism containing dozens of cell types that fight against threats both external (pathogens) and internal (cancer). These immune cells originate from hematopoietic stem cells (HSCs) and are composed of heterogenous populations that play distinct roles. Identifying *cis*-regulatory elements and gene networks orchestrating the immune response is a central goal of immunology. Transposable elements (TEs) are a source of species-specific regulatory novelty and have been shown to be co-opted as enhancers to regulate innate immunity genes. However, the *cis*-regulatory contribution of TEs across the immune system, including lymphocyte development, has not been comprehensively investigated. Mining more than 300 chromatin accessibility datasets (ATAC-seq and DNase-seq) from 10 immune cell types and 27 tissues from mouse, we identified rodent-specific TE families of retroviral origin called ORR1E and ORR1D2 (together termed ODE) that are exceptional for being enriched in accessible chromatin across immune cells and immune tissues. Interestingly, subsets of ~4,600 accessible ODEs exhibit distinct cell-type-specific accessibility patterns. ODEs among these subsets contain evidence of selection due to having significantly higher phastCons scores compared to nonspecifically accessible ODEs. Transcription factor (TF) motif enrichment analysis identified PU.1, a TF with critical roles in HSC differentiation and establishing lineage-specific identity of immune cells, as being broadly overrepresented across accessible ODEs. In addition, ODEs with cell-type-specific accessibility patterns contained enrichment for known immune cell-type-specific TFs such as EBF1, IRF8, and RORγ. These subsets of ODEs do not fall into phylogenetic clusters, suggesting the acquisition of discrete mutations post insertion to subfunctionalize their *cis*-regulatory activity. We correlated accessibility of ODEs with expression of nearby genes and found that ODEs putatively regulate hundreds of genes across discrete immune cell subsets. To elucidate the *cis*-regulatory potential of ODEs in a specific cell type, we identified enhancer-gene interactions by performing Micro-C in CD8+ T cells. We connected ~70 ODEs to ~110 target genes they may regulate and additionally linked ~240 ODEs to ~250 target genes using publicly available enhancer-promoter contacts in HSCs. Together, this work highlights the substantial contribution of ODEs to the regulation of the mouse immune system through their subfunctionalization as *cis*-regulatory elements across hematopoietic cell lineages.

EFFECTS OF INTRONIC L1 ORIENTATION ON HOST GENE EXPRESSION

Heather M Collazo[1], Tomoyuki Ohno[2], Partha S Saha[1], Yasunori Aizawa[2], Wenfeng An[1]

[1]South Dakota State University, Department of Pharmaceutical Sciences, Brookings, SD, [2]Tokyo Institute of Technology, School of Life Science and Technology, Yokohama, Japan

Mammalian genomes are littered with transposable elements, but their distribution is not uniform. In fact, sense-oriented L1s are 2-fold underrepresented in intronic locations relative to intergenic regions. It has been hypothesized that this orientation bias is the result of purifying selection against sense-oriented intronic L1s. Various mechanisms that can attenuate host gene expression have been proposed—including alternative splicing, inhibition of transcriptional elongation, and premature polyadenylation. However, the differential impact of sense and antisense L1 on gene expression has not been investigated from the same genomic locus.

Here, we are interested in understanding whether and how the orientation of an intronic L1 impacts host gene and its own transcription. To this end, we integrated a mouse L1 5'UTR-LacZ reporter construct into intron 1 of the ROSA26 gene locus in either sense or antisense orientation. When L1 promoter activity was examined, the antisense-oriented reporter showed much higher expression than the sense-oriented reporter across multiple tissues. Interestingly, when the host gene expression was examined, it was attenuated by both sense and antisense-oriented transgenes, although to a greater extent by the antisense-oriented transgene.

To check whether this phenomenon can be generalized, we biallelically knocked in a human L1 5'UTR-LacZ reporter construct into intron 1 of TP53 in either sense or antisense orientation in HCT116 cells. The parental alleles were already depleted of Alu sequences before our transgenes were introduced. Again, the antisense-oriented transgene showed higher expression than the sense-oriented counterpart. Like in the mouse model, the host gene was also downregulated much more prominently by the antisense-oriented transgene than by the sense-oriented transgene.

In summary, we recapitulated the crosstalk between host gene and L1 transgene in two different models. We plan to gain mechanistic insights by profiling the chromatin landscape across the targeted loci and perturbing specific pathways known to regulate intronic L1s. This work could contribute to our understanding of the orientation bias of intronic L1s and their differential effects on host gene expression.

RETROCOPIES ARE NUMEROUS AND WIDESPREAD IN PILOSA (XENARTHRA), A BASAL CLADE OF MAMMALS UNIQUE TO SOUTH AMERICA

Marcela Uliano-Silva[1,2], Helena B da Conceição[3], Rafael L. V. Mercuri[3,4], Mark Blaxter[1], Pedro A. F. Galante[3], Camila Mazzoni[5,6]

[1]Wellcome Sanger Institute, Tree of Life, Cambridge, United Kingdom, [2]Nord University, Faculty of Biosciences and Aquaculture, Bodø, Norway, [3]Hospital Sírio Libanês, Molecular Oncology Center, São Paulo, Brazil, [4]UNiversity of São Paulo, Interunits in Bioinformatics, São Paulo, Brazil, [5]IZW, Department of Evolutionary Genetics, Berlin, Germany, [6]Freie Universität Berlin, Begendiv, Berlin, Germany

Retrocopies are copies of mRNAs that are reverse-transcribed into the genome as a result of LINE1 activity. They are characterized by the conservation of only their parental exons, the frequent presence of poly(A) tail, and lack of their parental promoter regions. With the recent advancements in enabling multi-OMICs and bioinformatics technologies, an increasing number of retrocopies have been recognized as functional. Sloths, anteaters, and armadillos belong to the superorder Xenarthra, a basal clade of placental mammals unique to South America. We have sequenced chromosome-level genomes for the two-toed sloth Choloepus didactylus and the anteater Tamandua tetradactyla where 99.9% and 99.5% of DNA bases were assigned to 29 and 28 chromosomes, respectively. By applying our novel pipeline of retrocopy identification to these 2 genomes and 47 other chordate species (https://www.rcpediadb.org), we have identified that these Pilosa (Xenarthra) present an exacerbated number of retrocopies: 2.7 fold larger than in other mammals. This high number of retrocopies could indicate that the burst of retrotransposition was at the base of the Pilosa divergency. Curiously, only 10% of retrocopies are shared between C. didactylus and T. tetradactyla, indicating that most of these retrocopies events might have an independent origin. We have also identified retrocopies in the available genome of C. hoffmanni, showing that 60% of retrocopies are shared among them, evidencing their close evolutionary proximity. Additionally, our characterization of the mobile elements show that both Choloepus and T. tetradacyla genomes present recent insertions of LINE elements (less than 10% kimura divergency). We are further characterizing the recent LINE elements and their genome architectural and sequence evolution relationship with the retrocopies. We are also exploring signs of retrocopy expression and "domestication" in sloths and anteaters. These two new chromosome-level genomes for a genomically-neglected clade of placental mammals have enabled the characterization of a major burst of retrotransposition, including the origination of retrocopies, within this clade. Support: FAPESP 2024/00078-4, Horizon2020 Marie Skłodowska-Curie (750747)

EUKARYOTIC ALGAE AS SOURCES OF GIANT, EXOTIC AND ACTIVE MOBILE ELEMENTS

Rory J Craig[1], Eugenio Lopez-Cortegano[2]

[1]Max Planck Institute for Biology Tübingen, Algal Development and Evolution, Tübingen, Germany, [2]University of Edinburgh, Ecology and Evolution, Edinburgh, United Kingdom

Transposable elements (TEs) are selfish mobile genetic units that play fundamental roles in evolution, disease and biotechnology. TEs have evolved many times independently, display an incredible mechanistic diversity, and are near-ubiquitous across the tree of life. However, most of our understanding of eukaryotic TEs and host-TE evolutionary dynamics is derived from well-studied animal, plant and fungal genomes.

We will present results focussing on the diversity, activity, and evolution of transposons in algal genomes, with particular focus on *Chlamydomonas reinhardtii* (Chlamydomonas). The model green alga combines the experimental tractability of a microbial system with a relatively complex genomic architecture; the Chlamydomonas genome is ~114 Mb and has perhaps the most diverse TE repertoire of any model species. Via long- and short-read sequencing of experimental lines of multiple genotypes, we reveal a menagerie of actively transposing elements: the first active Cryptons, HUH transposons (Helitrons and Replitrons), and DD(E/D) DNA transposons from the poorly characterized *KDZ* and *Dada* superfamilies, in addition to active DIRS retrotransposons, *Penelope*-like elements and *Dualen* LINEs. Characterization of de novo insertions by genome assembly revealed several complex transposition properties, including a propensity for Cryptons to mediate genomic rearrangements and for *Dualens* to introduce large duplications up to 1.5 kb. Other active elements reach the threshold of giant transposons (>20 kb), such as a 20.4 kb Helitron and 31.7 kb *KDZ* DNA transposon. The heterogenous repertoire of active TEs in Chlamydomonas presents opportunities to functionally characterize understudied transposition mechanisms, with potential implications for biotechnology.

We will also present insights from TE manual curation in the genomes of Chlamydomonas and the model brown alga *Ectocarpus*, which collectively span billions of years of evolution. Both genomes feature giant DNA transposons, notably from the *KDZ* (green and brown algae) and *Plavaka* (brown algae) groups. These transposons regularly carry accessory genes, or even mobilize nested insertions of other TEs. Giant elements may be tolerated in many algal genomes due to life cycles with major haploid phases, where homologous recombination may be rare and the threat of ectopic recombination diminished.

A BUILT-IN FUNCTION OF THE FOUR-LETTER GENETIC ALPHABET – PATCHY SEQUENCE IDENTITIES BETWEEN GENOMES OF DIVERSE SPECIES ARE KEY TO ILLEGITIMATE RECOMBINATION, TRANSPOSON INSERTIONS AND EVOLUTION

Stefanie Weber[1], Christina M Ramirez-Kitchen[2], Walter Doerfler[1,3]

[1]Friedrich-Alexander University 91054 Erlangen, Germany, Virology, Erlangen, Germany, [2]UCLA School of Public Health, Los Angeles, CA 90095-1772, USA, Biostatistics, Los Angeles, CA, CA, [3]University of Cologne, Genetics, Cologne (Koeln), Germany

A comparison of nucleotide sequences between the genomes of Human Adenoviruses (Ad) and Severe-Acute-Respiratory-Coronavirus-2 (SARS-CoV-2) surprisingly revealed patchy sequence identity patterns of 42-48 % in which short stretches of identities alternated with those lacking them. In further investigations, sequences of completely unrelated species were aligned in over 60 analyses. The same 42-48 % of patch pattern identities were observed. Presumed “control” studies with differently shuffled sequences did not provide relevant information since they are subject to the same built-in mechanisms and, in fact, have generated the same patch type identities of ~ 45 % in comparative alignments. By an inherent mechanism, the four-letter genetic alphabet apparently allows to form patch type identities among genomes of completely unrelated species. Numerous publications report that comparable patchy sequence identities frequently occur at the junction sites of illegitimate recombination partners or of foreign DNA integrates. Such patterns might be recognized as signals for illegitimate recombination and hence could have played a relevant role in evolution. The four-letter alphabet may have been selected over innumerable cycles during evolution and prevailed ever since due to its ability to facilitate unlimited possibilities for illegitimate recombination and sequence exchange or integration reactions also for transposons. - A comparison of the Omicron BA.2.86 genome with that of the authentic SARS-CoV-2 Wuhan-Hu-1 also disclosed patches of 2 or 4 nucleotide pairs. Patch pattern recognition and illegitimate recombination might have contributed to the rapid rise of SARS-CoV-2 and its variants. Are there functions for site-selection in transposition events?
References: Weber S, Ramirez CM, Doerfler W. Viruses. 2022 Apr 24;14(5):885. doi: 10.3390/v14050885.

Weber S, Ramirez CM, Doerfler W. A built-in function of the four-letter genetic alphabet – patchy sequence identities between genomes of diverse species are key to illegitimate recombination and evolution. J. Mol. Biol., under revision, 2024.

Weber S, Ramirez CM, Weiser B, Burger H, Doerfler W. EMBO Mol. Med. 2021 Jun 7;13(6):e14062. doi:10.15252/emmm.202114062.

W. Doerfler. Clinical Epigenetics, Clinical Epigenetics (2020) 12:21. doi: https://doi.org/10.1186/s13148-020-0813-z -

TREAT OR TRICK--THE BATTLE BETWEEN TRANSPOSABLE ELEMENTS AND THE HOST

Dan Shen* [1], Yaqian Xu*[1], Qi Shi[1], Chongyang Li[1], Zhe Meng[1], Qiuju Wen[1], Chenhui Wang [1], Zhao Zhang[2,3], Kun Dou[1]

[1]ShanghaiTech University, School of Life Science and Technology, Shanghai, China, [2]Duke University School of Medicine, Department of Pharmacology and Cancer Biology, Durham, NC, [3]Duke University School of Medicine, Duke Regeneration Center, Durham, NC

Transposable elements represent the most abundant residents in the genome of nearly all eukaryotes, occupying half of the human genome and nearly twenty percent of the *Drosophila* genome. This indicates numerous rounds of invasions in the battle between transposable elements and the host during evolution. In these battles, transposons employ specific tactics for robust propagation. Previous studies showed that during *Drosophila* oogenesis, retrotransposons preferentially target the oocytes—the founder of the next generation--for propagation. But the underlying mechanism remains elusive. Here, by generating new tools, we try to unravel the tactics transposons utilize for robust propagation. We identified factors involved in transposon regulation, characterized their roles in the transposon-host battle, and raised up the model for transposon propagation. These unpublished data high light how transposons, especially specific retrotransposons, modulate the host developmental process for their own propagation. We will share our latest study on the transposon-host battle.

*These authors contributed equally to this work
Correspondence: doukun@shanghaitech.edu.cn

CHARACTERIZATION OF TRANSPOSON-ENCODED TNPB NUCLEASES

Gytis Druteika, Virginijus Siksnys, Tautvydas Karvelis

Vilnius University, Life Sciences Center, Institute of Biotechnology, Department of Protein-DNA Interactions, Vilnius, Lithuania

Intercellular DNA transfer plays a crucial role in driving evolution among prokaryotes. Mobile genetic elements (MGEs) such as insertion sequences (IS), plasmids, transposons, and phages are responsible for high genomic flexibility in bacteria, allowing rapid adaptation to varying environmental conditions [1]. Among these, the IS200/IS605 family of insertion sequences is one of the oldest and most widely distributed groups of MGEs in prokaryotes. These sequences are flanked by terminal imperfect palindromic motifs (LE and RE) and may encode either TnpA alone, TnpA and TnpB together, or just TnpB. While TnpA facilitates IS transposition, the role of TnpB has emerged recently [2, 3]. It was demonstrated that TnpB acts as an RNA-guided nuclease capable of cleaving double-stranded DNA substrates in a TAM-dependent manner. TnpB nucleases, being evolutionary ancestors of CRISPR-Cas12 effector proteins, were also applied for genome editing in human cell lines [3-5]. However, the requirement for TAM sequence limits the application of individual TnpB variants for genome editing. To address this issue, the characterization of novel TnpB variants remains of great interest.
Since TnpB-encoding insertion sequences are extremely widespread, this study aimed to characterize a set of TnpB orthologs. By combining in silico and biochemical analysis we were able to detect DNA cleavage activity and characterize diverse TAM sequences for TnpB proteins from the IS200/IS605 and IS607 families. Overall, these results contribute to the efforts to expand the genome editing toolbox with novel TnpB-based RNA-guided editors.

1. Krupovic, M. et al. Recent Mobility of Casposons, Self-Synthesizing Transposons at the Origin of the CRISPR-Cas Immunity. Genome Biol Evol 8, 375–386 (2016).
2. Meers, C., Le, H.C., Pesari, S.R. et al. Transposon-encoded nucleases use guide RNAs to promote their selfish spread. Nature 622, 863–871 (2023).
3. Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. Nature 599, 692–696 (2021).
4. Xiang, G. et al. Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. Nat Biotechnol 1–13 (2023).
5. Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. Science 374, 57–65 (2021).

THE EGG OR THE CHICKEN: STUDIES TO UNDERSTAND THE LINK BETWEEN TRANSPOSABLE ELEMENTS AND PD-NEUROINFLAMMATION

Raquel Garza[1], Anita Adami[1], Annabel Curle[2], Oliver Tam[3,4], Arun Thiruvalluvan[5], Talitha Forcier[3,4], Diahann Atacho[1], Annelies Quaegebeur[2], Cole Wunderlich[3,4], Agnete Kirkeby[5], Roger Barker[2], Molly Hammell[3,4], Johan Jakobsson[1]
[1]Laboratory of Molecular Neurogenetics, Department of Experimental Medical Science, Wallenberg Neuroscience Center and Lund Stem Cell Center, Lund University, Lund, Sweden, [2]Clinical Neurosciences, University of Cambridge and Department of Pathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom, [3]Institute for Systems Genetics, Department of Neuroscience and Physiology, NYU Langone Health, New York, NY, [4]Neuroscience Institute, NYU Grossman School of Medicine, New York, NY, [5]Novo Nordisk Foundation Center for Stem Cell Medicine, Department of Neuroscience, Copenhagen, Denmark

Neuroinflammation is a hallmark of Parkinson's disease (PD), yet the trigger and progression of it is not understood. We performed single-nuclei RNAseq on post-mortem brain tissues from PD patients and control individuals (n=150) where we observed a **region-specific neuroinflammatory response in PD that correlates with an increased expression of** transposable elements (TE), such as human **endogenous retroviruses (HERV)**. HERV elements have been previously implicated with the beginning of human neuroinflammation. However, their specific role in the neuroinflammatory process is not understood and **has not been previously documented in PD**. To better understand HERVs spatial distribution in relation to the disease progression, we are now extending our investigation using spatial data of custom neuroinflammatory-genes and TE probes. However, key questions remain to be **does the aberrant transcriptional activation of HERV lead to an inflammatory response?** And **does HERV transcription increase as a result of an inflammatory state?** Thus, to mechanistically link the transcriptional activation of HERVs and human neuroinflammation, we performed a series of in-vitro experiments. To study the molecular consequences of an aberrant HERV expression, we developed a CRISPR-activation system (CRISPRa) targeting these elements. CRISPRa vectors were tested in human neural progenitor cells (hNPCs), which resulted in robust transcriptional activation of dozens of HERV elements. Our results show that the **activation of HERV in hNPCs results in the upregulation of stress-response and inflammatory-related genes.** Moreover, when **introducing pro-inflammatory cytokines to microglia and astrocyte cultures, we observe a robust transcriptional activation of HERV elements**. Our studies suggest HERV elements have a role in PD-neuroinflammation and mechanistically link their aberrant transcription to the beginning of human neuroinflammation.

A COMPARATIVE ROADMAP OF PIWI-INTERACTING RNAs (piRNAs) ACROSS SEVEN SPECIES REVEALS INSIGHTS INTO DE NOVO piRNA PRECURSOR FORMATION IN MAMMALS.

Parthena Konstantinidou*[1], Zuzana Loubalova*[1], Franziska Ahrend*[1], Aleksandr Friman*[1,5], Miguel Vasconcelos Almeida[2], Axel Poulet[3], Filip Horvat[4], Yuejun Wang[1], Wolfgang Losert[5], Hernan Lorenzi[1], Petr Svoboda[4], Eric A Miska[2], Josien C van Wolfswinkel[3], Astrid D Haase[1]

[1]National Institutes of Health, National Institutes of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, [2]University of Cambridge, Department of Biochemistry, Cambridge, United Kingdom, [3]Yale University, Department of Molecular Cellular and Developmental Biology, New Haven, CT, [4]Czech Academy of Sciences, Institute of Molecular Genetics, Prague, Czech Republic, [5]University of Maryland, Department of Physics, College Park, MD

PIWI-interacting RNAs (piRNAs) play a crucial role in safeguarding genome integrity by silencing mobile genetic elements. From flies to human, piRNAs originate from long single-stranded precursors encoded by genomic piRNA clusters. How piRNA clusters form to adapt to novel genomic invaders and evolve to maintain protection, remain key outstanding questions. Here, we generated a roadmap of piRNA clusters across seven species that highlights both similarities and variations. In mammals, we identified transcriptional readthrough as a mechanism to generate piRNAs from transposon insertions downstream-of-genes (piC-DoGs). Together with the well-known stress-dependent DoG-transcripts, our findings suggest a molecular mechanism for the formation of piRNA clusters in response to retroviral invasion. Finally, we identified a novel class of dynamic piRNA clusters in humans, underscoring unique features of human germ cell biology. Our results advance the understanding of conserved principles and species-specific variations in piRNA biology and provide tools for future studies.

COMPREHENSIVE MAPPING OF ENDOGENOUS L1 INSERTIONS FROM MOUSE TISSUE OR SINGLE CELLS

Xuanming Zhang, Ivana Celic, Hannah Mitchell, Sam Stuckert, Lalitha Vedula, Jeffrey Han

Tulane University School of Medicine, Biochemistry and Molecular Biology, New Orleans, LA

The association of L1 expression with many health disorders in mammals has raised interest in the mapping of *de novo* insertion activity of endogenous L1 elements *in vivo*, particularly during the progression of disease. A variety of useful methods have been developed for the mapping of endogenous human L1s from clinical samples. Adapting these methods to mouse genomes is desirable because it would allow us to monitor endogenous in vivo L1 activity in controlled, experimental conditions using mouse disease models. However, adapting these methods to mouse genomes presents additional challenges. The repertoire of active L1s is more complex in mouse than human, with multiple active subfamilies and much greater number of potentially active elements. We modified a previously described transposon insertion profiling-seq (TIPseq) protocol to selectively enrich the young, potentially active mouse L1s responsible for most naturally occurring mouse L1 mutations. By linking this amplification step with long-read Oxford nanopore sequencing (nanoTIPseq), we were able to identify virtually all relevant annotated L1s from C57BL/6 genomic DNA with only 200,000 sequencing reads. We also identified 82 unnannotated L1 insertions from a single C57BL/6 genome. Most of these polymorphic L1s were near repetitive sequence and not found with short-read TIPseq. We used nanoTIPseq on whole genome amplified 4226 single cells (a line derived from a MMTV-Wnt1 cancer model) and were able to reproducibly identify unannotated L1s specific to this cell line. Here we will further discuss our efforts to monitor retrotransposition in the germ line, by applying nanoTIPseq to single cells from various stages of spermatogenesis.

TRANSCRIPTIONAL DYNAMICS OF L1 ANTISENSE PROMOTERS DURING DIFFERENTIATION OF HUMAN iPS CELLS INTO β-cells

Hinata Hashimoto[1], Kengo Matsuzawa[1], Nobuaki Shiraki[1], Yasunori Aizawa[1,2,3]

[1]Tokyo Institute of Technology, Life Science and Technology, Kanagawa, Japan, [2]Kanagawa Institute of Industrial Science and Technology, Synthetic Biology, Kanagawa, Japan, [3]Logomix,, Tokyo, Japan

Long interspersed element 1 (L1) retrotransposons are widely distributed in the human genome, occupying approximately 500,000 loci. Of them, ~10,000 loci maintain the full-length copies. Although the majority of L1 elements have lost their retrotransposition ability, a significant subset of human L1 elements maintains bidirectional transcriptional activity from their internal promoters. The sense promoter drives the transcription of the entire L1 mRNA, leading to L1 mobilization, while the antisense promoter (ASP) transcribes L1 chimeric RNAs that include adjacent exon sequences(1). Our group previously developed a quantitative method to identify genomic loci with high L1 ASP activity and applied this technique called LATRAP (L1 antisense transcriptome protocol) to cancer cells(2). In cancer cells, the activity of L1 ASP has been shown to significantly contribute to gene expression(3). However, activation status of L1 ASP in normal cellular differentiation processes remains poorly investigated.
In this study, we applied LATRAP to perform genome-wide identification of L1 loci that were transcriptionally up- and down-regulated during the differentiation of human iPS cells into mature β-cells. Comparison with conventional RNA-seq data at the same differentiation stages uncovers the presence of L1 elements that are co-regulated with the expression changes of surrounding genes, as well as those that exhibit expression changes completely independent of them. In this presentation, we will discuss biological significance of L1 ASP activity during the differentiation process.

(1)Raquel Garza *et al.*, LINE-1 retrotransposons drive human neuronal transcriptome complexity and functional diversification. *Sci. Adv.* 9, eadh9543 (2023)
(2)Koichi Ishiguro *et al.*, Establishment of a genome-wide and quantitative protocol for assessment of transcriptional activity at human retrotransposon L1 antisense promoters. *Genes Genet. Syst.*, 92, 5, 243-249(2017)
(3)Jose M C Tubio *et al.*, Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science*. Aug 1, 345 (2014)

GENDER AND HISTOLOGIC SPECIFIC DIFFERENCES IN URINARY EXOSOMAL LINE-1 AND miRNA PROFILES AS DIAGNOSTIC BIOMARKERS OF NON-SMALL CELL LUNG CANCER

Abeer A Hassanin[1,2], Kenneth S Ramos[1]

[1]Texas A&M Institute of Biosciences and Technology, Texas Medical Center, Center for Genomic and Precision Medicine, Houston, TX, [2]Faculty of Veterinary Medicine, Suez Canal University, Department of Animal Wealth Development, Ismailia, Egypt

Lung cancer is the main cause of cancer-related deaths in both males and females globally. The optimal treatment approach for squamous cell carcinoma (SQCLC) and adenocarcinoma (LUAD) has benefited from precision-based approaches and therefore, there is an urgent demand for non- invasive biomarkers that facilitate an accurate diagnosis and inform clinical management. In a previous study, we assessed the utility of plasma exosomal Long Interspersed Element-1 (LINE-1) analytes as diagnostic and prognostic biomarkers in non-small cell lung cancer (NSCLC) patients. In the present study we extend our efforts by examining patterns of LINE-1 expression to urine-derived exosomes of patients with NSCLC. Exosomal levels of LINE-1 ORF1 and ORF2 mRNAs and nine related-miRNAs were examined in urine samples of ostensibly healthy controls and patients with SQCLC or LUAD. We report that LINE-1 mRNAs and a set of microRNAs, including miR-21-5p, miR-126-3p, miR-210-3p, miR-221-3p, Let-7b-5p, miR-146a-5p, miR-222-3p, miR-9-5p, and miR-1277-5p, were readily detected in both female and male patients. The mRNAs of both LINE-1 ORF1 and ORF2 were higher in exosomes from female patients compared to males in both lung cancer subtypes, and in metastatic tumors with a size T ≥3. Exosomes from male and female patients with SQCLC and LUAD patients also exhibited significantly higher levels of all miRNAs examined, with patterns of expression that correlated with the clinicopathological characteristics of tumors. ROC curve analysis demonstrated that exosomal LINE-1 in females was superior in distinguishing SQCLC and LUAD patients, as well as differentiating between the two histological subtypes (AUC =0.894 for SQCLC, 0.969 for LUAD, and 0.988 for SQCLC vs LUAD), compared to their male counterparts (AUC =0.669 for SQCLC, 0.823 for LUAD, and 0.846 for SQCLC vs LUAD).Together, these findings suggest that urinary exosomal LINE-1 analytes may serve as diagnostic and prognostic biomarkers of NSCLC. Because the expression levels of urinary exosomal analytes were comparable to those previously detected in plasma, their utility as a readily accessible biofluid may offer additional options in the evaluation of patients with NSCLC.

ROLE OF TRANSPOSABLE ELEMENTS IN A PARASITOID WASP ONGOING SPECIATION

Camille Heisserer[1], Mohamed Amine Chebbi[1], Emilie Brasset[2], Silke Jensen[2], Elisabeth Huguet[1], Karen Kester[3], Thibaut Josse[1], Jean-Michel Drezen[1]

[1]Université de Tours, CNRS, Institut de Recherche sur la Biologie de l'Insecte, Tours, France, [2]Université Clermont Auvergne, CNRS, Génétique, Reproduction et Développement, Clermont-Ferrand, France, [3]Virginia Commonwealth University, Department of Biology, Richmond, VA

Endoparasitoid wasps lay their eggs and develop at larval stages inside a specific host. Recent studies have shown that endoparasitoid wasp species often constitute a complex of sister species, each one resulting from the adaptation to a particular host. This project aims to understand the involvement of transposable elements (TE) in the ongoing speciation of two populations of the wasp *Cotesia congregata* (CcC and MsT) specialized on different hosts and with two distinct ecological niches. The CcC population parasitizes the caterpillar *Ceratomia catalpae* which lives on the catalpa tree, and MsT population parasitizes the caterpillar *Manduca sexta* which lives on tobacco. Differences in reproductive behavior and genetic differentiation (microsatellites and COI) indicate these two populations are beginning a speciation process.

Interestingly, a reproductive defect is observed which could contribute to reinforcing barriers between the two populations. The cross between CcC female and MsT male gives a fertile offspring while the reciprocal cross (MsT female X CcC male) gives a nearly sterile offspring showing ovaries atrophy. We hypothesize that this phenotype corresponds to hybrid dysgenesis previously described in *Drosophila*, which would be induced by a TE present and active in CcC and not in MsT wasps or a more global deregulation of TE control. Hence, the TE could contribute to speciation by limiting genetic exchanges between the two populations. This study is the opportunity to study a TE-invading wasp population whereas P and I elements were studied in *Drosophila* long after their genomes had been invaded worldwide.

Here, we present results identifying TE candidates potentially involved in dysgenesis among those showing piRNA repression in CcC and not in MsT. We present a study of ping-pong repression against TE thanks to piRNA sequencing of samples collected over several years. We also studied ping-pong repression in the head in comparison to repression in the gonads. Finally, we characterize and compare piRNA clusters of the two wasp populations.

THE STUDY OF HISTONE MODIFICATIONS INDUCED BY *ACK* sRNAs ON *AC* ELEMENTS

Lilly Horowitz[1], Dafang Wang[2]

[1]Jericho High School, Jericho, NY, [2]Hofstra University, Biology, Hempstead, NY

Ac/Ds transposons, the first discovered transposable elements, are Type II transposons. *Ac* killer (*Ack*), a trans-acting silencer generated by Ac alternative transposition, produces 21/22 nt and 24 nt small RNAs with sequence homology to Ac elements in maize. These small RNAs can target and initiate silencing of active Ac elements, though the exact silencing mechanism remains unclear. While it has been determined that these small RNAs can initiate DNA methylation and silence *Ac* elements, their relationship to histone modifications in respect to silencing Ac elements is unknown. In this study, we examined the effects of these novel small RNAs on two histone modification marks (H3K9me2 and H3K27me3) across four developmental stages: germinating embryo, juvenile leaf (leaf 3), adult leaf (leaf 10), and silk. We observed varied enrichment of these histone modifications across different developmental stages at areas targetted by *Ack* small RNAs and those that are unaffected. Differences in histone modifications may explain the decreased *Ac* transcription and *Ac* activity detected at each stage.

NON-CELL AUTONOMOUS TOXICITY AND MDG4 ENDOGENOUS RETROVIRUS REPLICATION IN A *DROSOPHILA* MODEL OF C9ORF72 ALS/FTD

Isabel Hubbard[1], Josh Dubnau[1,2]

[1]Stony Brook University, Neurobiology and Behavior, Stony Brook, NY,
[2]Stony Brook Medicine, Anesthesiology, Stony Brook, NY

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are fatal neurodegenerative diseases that lead to severe cognitive and motor defects in patients and both feature aggregation of the protein TDP-43. Non-cell autonomous effects are a hallmark of these disorders, which we have previously reported to be mediated at least in part by endogenous retroviruses. The most common genetic cause of both familial ALS and FTD is a hexanucleotide repeat sequence consisting of four guanine and two cytosine nucleotides (G_4C_2) in the first intron of the gene C9orf72, which can produce expansions of up to hundreds to thousands of repeats in patients. The C9orf72 repeat expansion is bidirectionally transcribed into sense and anti-sense RNA foci. Additionally, a non-canonical translation mechanism known as repeat-associated non-ATG translation (RAN translation) allows the expansion to be translated into five different dipeptide repeats (DPRs): GA, GP, GR, PA, and PR. Previous work has shown that multiple forms of the repeat cause premature death and neurodegenerative effects in animal models, including *Drosophila*, when expressed neuronally. The arginine-rich DPRs produce the highest level of toxicity. Effects on glia, impacts on non-cell autonomous toxicity, and on activation of retrotransposons of the DPRs and RNA repeats each are understudied. Here, we use glial cell type-specific expression of individual DPRs, of RNA repeat-only, or of the G_4C_2 repeat that is capable of producing both DPRs and RNA repeats, to systematically investigate both cell-intrinsic and non-cell autonomous toxicity of each of these components. With pan-glial expression, the G_4C_2 repeat and GR dipeptide are equally the most toxic forms of the repeat; but when focally expressed in subperineural glia (SPG), a glial subtype, the pure G_4C_2 repeat is more toxic to the animal than is the GR dipeptide. Imaging experiments reveal that while both G_4C_2 and GR produce similar levels of toxicity to the cells in which they are expressed in, the pure G_4C_2 repeat causes the greatest degree of non-cell autonomous toxicity to other glial subtypes and neurons. The G_4C_2 and GR repeats, namely those capable of producing the GR DPR, also each triggered replication of the fly endogenous retrovirus mdg4, which has been shown to mediate non-cell autonomy in TDP-43 fly models. Expression of the apoptosis inhibitor gene p35 shortened lifespans in flies focally expressing G4C2 and GR in SPG, suggesting that keeping the cells expressing these both of these repeats alive longer increases non-cell autonomous toxicity.

KRAB ZINC FINGER PROTEINS EXPAND THE DOMESTICATION POTENTIAL OF MOBILE ELEMENTS.

Athina Triantou, Samuel Daffern, Juliette Davis, Denali Archer, Santiago Morell, Michael Imbeault

University of Cambridge, Department of Genetics, Cambridge, United Kingdom

KRAB zinc finger proteins (KZFPs) are the largest DNA binding protein subfamily in tetrapods. Their distinctive architecture combines a zinc finger array and a KRAB domain, enabling them to precisely target specific DNA sequences where they induce the formation of heterochromatin through interaction with TRIM28. KZFPs have a remarkable affinity for transposable elements (TEs) and contribute to their transcriptional silencing, restricting their transposition potential. Interestingly, while arms race dynamics with TEs are hypothesized to fuel the rapid evolution of KZFPs, most human KZFPs were found to be evolutionary conserved and bind to equally ancient TE remnants, suggesting that they participate in the epigenetic control of domesticated regulatory platforms. I will report on our recent research in elucidating the role of select KRAB zinc finger proteins, encompassing their engagement with TEs and their influence on important biological pathways.

I will showcase how primate specific ZNF808 is essential for pancreatic development and a rare cause of neonatal diabetes in humans. We found that in its absence during early development the fate of pancreatic progenitors is incorrectly skewed toward the hepatic lineage, in part due to loss of epigenetic silencing on MER11 endogenous retroelements. These are enriched in transcription factor binding sites of the GATA and HNF families relevant for development of diverse tissues and organs. We further demonstrate that MER11 elements are normally active and potentially domesticated in select cellular contexts where ZNF808 is lowly expressed, notably placenta and liver. This suggests that KZFPs expand the domestication potential of TEs by allowing individual elements to be active where and when they are potentially beneficial while restricting them in cellular contexts where they are not.

Moving beyond development, I will highlight how KZFPs can dynamically regulate aspects of the response to pathogens or DNA damage in humans. We found a subset of primate KZFPs that are inducible by specific environmental stimuli (either inflammation or genotoxic stress) and target TEs that are highly enriched in transcription factor binding sites relevant for the same stimulus. We show evidence that KZFPs and TEs can combine to form dynamic circuits of gene regulation that can drive complex time-dependent expression programs in response to environmental cues. These few examples show that KZFPs have the potential to orchestrate complex regulation at the interface between TEs, developmental processes and biological responses relevant to human health.

THE EFFECT OF *MOP1* MUTATION ON EPIGENETIC SILENCING OF *AC* INDUCED BY *AC KILLER* SRNAS

Zuhair Ishraq[1], Charlotte Huang[2], Dafang Wang[1]

[1]Hofstra University, Biology Department, Hempstead, NY, [2]W Tresper Clarke High School, Biology Department, Westbury, NY

Maize *Ac/Ds* transposable elements, classified as Class II transposons, were the first transposons discovered by Barbara McClintock. In a recent study, a naturally occurring silencer, *Ac killer* (*Ack*), was characterized, which is derived from an alternative transposition event of *Ac*. This silencer produces novel 21/22 nt and 24 nt small RNAs from *Ac* sequences, allowing them to target active *Ac* elements. Preliminary data show that small RNAs from *Ack* induce RNA-directed DNA methylation in all contexts and histone modifications. However, it remains unclear which classes of small RNA are the primary contributors.

In this study, we hypothesize that specific classes of siRNAs are responsible for these epigenetic changes. We take advantage of a loss-of-function mutation in the mediator of paramutation1 (mop1), which encodes RDR2 in the wild type and significantly reduces the accumulation of 24 nt small RNAs in the mutant. This mutation allows us to alter the profile of small RNAs from *Ack*. By studying the new DNA methylation profile and histone modifications in the maize vegetative stage, we aim to demonstrate the distinct contributions of 21/22 nt and 24 nt small RNAs to *Ack*-induced epigenetic silencing.

ENDOGENOUS RETROVIRUSES ARE A SOURCE OF REGULATORY VARIANTS IN COLORECTAL CANCER

Atma Ivancevic, Edward Chuong

University of Colorado Boulder, Department of Molecular, Cellular & Developmental Biology, Boulder, CO

Transposable elements such as endogenous retroviruses shape chromatin landscapes. They facilitate the generation of novel cell-type-specific regulatory networks by replicating themselves and dispersing hundreds of potential regulatory sequences throughout the genome. Despite their recognized importance in cell biology, repetitive sequences continue to be understudied, especially as a source of regulatory elements that are epigenetically silenced in healthy cells but become activated in diseased cells.

Recently, we used multi-omic sequencing data and CRISPR experiments to show that ancient LTR10 elements can reactivate as tumor-specific enhancers in human colorectal cancer cells. LTR10 elements are derived from the long terminal repeat of a primate gammaretrovirus, HERV-I, which integrated into the anthropoid genome 30 million years ago. While heavily degraded in the human genome, many LTR10 sequences still contain binding motifs for the AP1 transcription factor complex. We found that LTR10 elements causally drive AP1-dependent expression of genes with established roles in tumorigenesis and therapy resistance, including ATG12, XRCC4, and VCAN. Bulk and single cell RNA-seq from patient tumors showed that LTR10 elements are transcriptionally activated in a substantial fraction (~30%) of cases.

We further discovered that individual LTR10 sequences contain variable number tandem repeat regions that affect AP1 binding activity. Although all LTR10 insertions are fixed in the human population, the number of AP1 motifs within an LTR10 sequence varies between individuals and even between tumors. We observed extensive germline genetic variation at AP1 motifs within LTR10 elements, supported by both short-read and long-read datasets. Additionally, long-read whole-genome sequencing from matched colorectal tumor and normal tissues revealed evidence of tumor-specific LTR10 expansions, sometimes thousands of bp in length. We suspect these LTR10 expansions drive tumor-specific gene regulatory activity. Altogether, our work implicates endogenous retroviruses like LTR10 elements as sources of pathological regulatory variants that facilitate transcriptional rewiring in cancer.

GENOMIC INTEGRATION OF SARS-COV-2 SEQUENCES IN VIRUS INFECTED AND VIRAL RNA TRANSFECTED HUMAN CELLS

Qi Liu, Rudolf Jaenisch

Whitehead Institute, Biology, MIT, Cambridge, MA

Background: SARS-CoV-2 sequences can be reverse-transcribed and integrated into the genomes of virus-infected cells by a LINE1-mediated retrotransposition mechanism. Whole genome sequencing (WGS) methods revealed the footprints of LINE1 mediated retrotransposition (target site duplication, presence of a LINE1 endonuclease recognition consensus sequence and presence of a polyadenylation sequence at the integration site) indicating integration by the well-defined "target-site primed reverse transcription" mechanism (1). We have recently confirmed these results (2). Indirect evidence also indicates that retrotransposition of viral sequences occurs in tissues of virus infected patients. In my talk I will summarize new results that extend our published work.

A. Isolation of cells carrying a retrotransposed SARS-CoV-2 sequences. One of the major limitations to study the consequence of retrotransposed viral sequences is that viral infection is toxic causing cell death within 2 to 3 days thus preventing the isolation of cells carrying integrated SARS-CoV-2 sequences. To overcome this limitation, we have developed a reporter system using subgenomic nucleocapsid RNA for transfection (the same RNA as synthesized in infected cells). This allows isolation of cells carrying a single retrotransposed copy and to assess the spectrum of integrations.

B. Does vaccine RNA integrate? This is a question of considerable public interest as potential integration of vaccine RNA into the genome could lead to mutations and is used as argument against vaccination. We have shown previously that nuclear capsid sub-genomic RNA transfected into cells does not integrate into the genome in contrast to viral RNA in virus infected cells (2). Using the RNA reporter assay described above we now find that the NC reporter RNA can retrotranspose when transfected into cells that over-express LINE1 (positive control). We are currently injecting the reporter RNA into mice to assess whether retrotransposition can occur in vivo.

1. Zhang L, Richards A, Barrassa MI, Hughes SH, Young RA and Jaenisch R. Reverse-
transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and
can be expressed in patient-derived tissues. Proc. Nat. Acad. Sci. USA 118 (21) e2105968118. (2021). PMCID: PMC8166107

2. Zhang L, Punam B, Flamier A, Barrasa I, Richards A, Hughes S and Jaenisch R. LINE1- mediated reverse transcription and genomic integration of SARS-CoV-2 mRNA detected in virus-infected but not in viral mRNA-transfected cells, Viruses 15, 629, 10.3390/v15030629 (2023)

A PHYLOGENETIC APPROACH FOR ESTIMATING THE AGE OF INDIVIDUAL TRANSPOSABLE ELEMENTS IN MAMMALIAN GENOMES

Pakkanan Chansongkrow, Diana Voicu, Jeremi Jaksina, Michael Imbeault

University of Cambridge, Department of Genetics, Cambridge, United Kingdom

Transposable elements (TEs) have been integrating into mammalian genomes and shaping them ever since the inception of the clade, over 100 million years ago. Each species harbours its unique combination of TEs acquired over the course of its evolution, therefore containing a mix of both species-specific and ancestrally-shared elements. While it is trivial to obtain a relative ranking of age at the subfamily level within a species, estimating the age of individual transposable elements is challenging and requires making assumptions regarding mutation rates and selection pressures over a long period of time. Here we present a novel method for estimating the age of transposons based on phylogenetic analysis and evolutionary conservation, leveraging the Zoonomia multiple sequence alignment (MSA) of 241 high quality mammalian genomes. To estimate the age of individual integrants, we test for its presence in other species at the same syntenic locus,, and also verify the presence of 5 kb flanks for each side. This way, we can assign each transposon to a point of divergence with the oldest species within the alignment that shows a significant homology at the same orthologous site. Alignments without flanking matches in any species were tagged as undefined, as they are potentially species-specific genomic rearrangements. As a result, we have tracked the origins of every TE subfamily at the individual integrant level in both the human and mouse genomes. This approach overcomes the limitations of previous techniques by making no additional assumptions about the mutation rates or generation times. Our pipeline can be used for any mammal within the Zoonomia dataset or any set of species for which an appropriate whole genome multiple alignment is available. Overall, this method reveals a more precise timeline of TE integration and activity across human and murine evolution. The phylogenetic approach offers a robust framework for future studies on the evolutionary history and functional impact of transposable elements.

CREATION OF A UNIFIED BENCHMARK STRATEGY FOR AUTOMATED TRANSPOSABLE ELEMENT DISCOVERY PIPELINES

Hannah P Kania, Sierra Seifert, Anne D Yoder

Duke University, Department of Biology, Durham, NC

Transposable elements (TEs) are mobile genetic units that comprise at least 45% of the human genome. In fact, TEs are almost ubiquitous across genomes spanning the tree of life, where they influence genome evolution, gene expression, and disease onset. However, due to their complex replication patterns, TEs are challenging to find and classify. Automated discovery pipelines attempt to solve this problem, but there is not yet a standard benchmark to evaluate their performance. Our study aims to create this benchmark and use it to compare three popular and promising discovery pipelines: RepeatModeler2, EDTA, and EarlGrey.

Our study benchmarks the discovery pipelines using simulated sequences built under different taxonomic contexts. We simulated sequences using Garlic, a program which leverages curated annotations of genes and repeat regions of a specified genome to produce artificial sequences that mimic intergenic regions. The artificial sequences are representative of the complex landscape of intergenic GC content and repeat regions across the genome, as opposed to only considering summary values. When prompted, Garlic can also generate background sequences with 0% repeat content, which serve as negative controls. Since all repeats, or absence of them, are known, the simulated sequences enable benchmarking of each discovery pipeline down to individual base pairs and TEs, extending beyond the TE-family level.

To build genome models, we used curated annotations from the University of California Santa Cruz Genome Browser and associated genome hubs. We produced 100Mb pseudo-sequences for five species representing taxonomic and repeat diversity: human, megabat, fruit fly, baker's yeast, and rice. For each species, a background sequence without repeat content and a sequence mimicking species-specific intergenic attributes were generated. Furthermore, we took advantage of existing long-read and short-read data from *Drosophila melanogaster* to build two genome assemblies around highly-curated repeat regions. The two assemblies purposely differ in input sequencing read length, as we aimed to gain a benchmark that describes the ability for each pipeline to identify TEs under varying sequencing contexts. After running the artificial sequences and *D. melanogaster* assemblies through each discovery pipeline, we benchmarked the resulting TE annotations against the established repeat content. We calculated metrics for precision, accuracy, sensitivity, specificity, and false discovery rate down to individual base pairs and up to the level of TE-families. Our poster will present preliminary results from our benchmark study and discuss ongoing analyses.

FASCINATING INSTANCES OF HORIZONTAL TRANSFER OF MARINER TRANSPOSONS

Vladimir V Kapitonov

Marine Biological Laboratory, Josephine Bay Paul Center, Woods Hole, MA

Here I report two instances of horizontal transfer (HT) involved in evolution of Mariner transposons. HSMAR2 is a prominent family of Mariner transposons that was active about 60 million years ago (MYA) in primates. Its antiquity is evidenced by 86% identity of the HSMAR2 consensus sequence to its copies present in the human genome. Surprisingly, HSMAR2 is also present in the genome of *Darwinula stevensoni* (Crustacea; Ostracoda)[1], which separated from humans 700 MYA. This assembly contains >10 copies of HSMAR2_DSte. The HSMAR2_DSte consensus sequence is 93% identical to human HSMAR2. This transposon was active in *D. stevensoni* as recently as 1-5 MYA, given the 97% identity of its copies to their consensus sequence. This observation raises an interesting paradox: why the HSMAR2 and HSMAR2_DSte transposons are 93% identical to each other despite nearly 60-million-year interval separating their activity peaks. So far, the most parsimonious explanation is that HSMAR2 existed/exists in genome(s) of not yet sequenced species as a domesticated functional element under strong negative selection and actively transposing at the same time.

The second case of HT was found also in the *D. stevensoni* assembly, containing >10 copies of Mariner-HT1_DSte that are 96% identical to their consensus sequence. This consensus is 96% identical to the HSMAR-N2_PSi (Repbase), present in the *Pelodiscus sinensis* turtle genome. HSMAR-N2_PSi is a 953-bp non-autonomous Mariner transposon that was active >58 MYA in the common ancestor of Trionychidae (soft-shelled turtles; genomes of 4 different species). The ancient ancestral autonomous element involved in transposition of HSMAR-N1 >58 MYA is not found in Trionychidae assemblies. However, it can be reliably approximated by the consensus sequence of HSMAR-HT1_DSte that was active recently (1-5 MYA) in *D. stevensoni*, separated from turtles by 700 million years of evolution.

[1] Tran Van P, Anselmetti Y, Bast J, Dumas Z, Galtier N, Jaron KS, Martens K, Parker DJ, Robinson-Rechavi M, Schwander T, Simion P, Schön I. (2021) First annotated draft genomes of nonmarine ostracods (Ostracoda, Crustacea) with different reproductive modes. G3 (Bethesda), 11(4):jkab043.

IDENTIFICATION AND GENOMIC NEIGHBORHOOD ANALYSIS OF IMMOBILIZED TRANSPOSABLE ELEMENTS

Alon Katz, Yuri I Wolf, Kira S Makarova, Eugene V Koonin

National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information, Bethesda, MD

Transposable elements (TEs) are among the most abundant genomic features in nature and are considered major drivers of evolution. Some TEs have lost the transposition capacity and were 'domesticated', acquiring a function in the host cell, such as defense against other mobile genetic elements as exemplified by type II and type V CRISPR-Cas systems. We performed a large-scale analysis of prokaryotic genomes to identify clades of immobilized TEs and investigate conserved elements in their genome neighborhoods. TEs were clustered into 45 alignable clusters (operationally, "families"). To estimate mobility, we counted the number of nearly identical transposases in each completely sequenced prokaryote genome; a 98% identity cluster was considered immobile if it was not found in more than one copy per genome. A phylogenetic tree was reconstructed for each TE family, and tight clades of immobile elements were identified. The local genome contexts of each TE gene from the immobile clades were explored to assess the strength of the association between the TE and its neighboring genes. Overall, 1,122 immobile clades in 40 TE families were identified. The families with the largest fraction of immobile TEs include RNA-guided transposable nucleases of the IscB (97%) and TnpB (90%) families, but also serine integrase (94%) and MuA transposase associated with both Mu and Tn7 TE families (93%). Analysis of gene content of "immobile" clades reproduced many conserved associations identified previously in the TnpB family, but also revealed dozens new conserved associations for other TE families.

A TRANSPOSASE-DERIVED GENE REQUIRED FOR HUMAN BRAIN DEVELOPMENT

Alex Kentsis

Memorial Sloan Kettering Cancer Center, Sloan Kettering Institute, New York, NY

DNA transposable elements and transposase-derived genes are present in most living organisms, including vertebrates, but their function is largely unknown. PiggyBac Transposable Element Derived 5 (PGBD5) is an evolutionarily conserved vertebrate DNA transposase-derived gene with retained nuclease activity in cells. Vertebrate brain development is known to be associated with prominent neuronal cell death and DNA breaks, but their causes and functions are not well understood. Here, we show that PGBD5 contributes to normal brain development in mice and humans, where its deficiency causes disorder of intellectual disability, movement, and seizures. In mice, Pgbd5 is required for the developmental induction of post-mitotic DNA breaks and recurrent somatic genome rearrangements. In the mouse cortex, loss of Pgbd5 leads to aberrant differentiation and gene expression of distinct neuronal populations, including multiple types of glutamatergic neurons, potentially explaining the features of PGBD5 deficiency in humans. Together, these studies nominate PGBD5 as the long-hypothesized neuronal DNA nuclease and transposase-derived gene required for brain function in mammals.

CHARACTERIZATION OF L1 ENDONUCLEASE INTERACTIONS WITH SMALL MOLECULE OR DNA SUBSTRATES

Hanna M Kodama, Alexandra D'Ordine, Martin Taylor, Gerwald Jogl, John Sedivy

Brown University, Molecular Biology, Cell Biology, and Biochemistry, Providence, RI

Long Interspersed Nuclear Element 1 (LINE1) is the only autonomous retrotransposon in humans, relying on a copy-and-paste mechanism which has led it to comprise 17% of the genome. This parasitic gene element is linked to cancer, aging, and neurodegenerative diseases due to its activity rearranging genes and triggering inflammatory cellular responses. It encodes two proteins, ORF1 and ORF2, which together mobilize and complete the target-site primed reverse transcription of LINE1 RNA. The endonuclease (EN) domain of ORF2 performs the first enzymatic action at the site of insertion, nicking the genomic DNA at a 5'-TTTTTAA-3' consensus sequence. Recent work shows that the enzyme displays in vitro target specificity deviating from the canonical consensus sequence and suggests the mechanism relies more heavily on structural inputs instead. Here, we employ biochemical and biophysical techniques to examine the activity and binding of EN with an array of substrates selected to explore the scope of its capabilities in recognizing nucleic acid secondary structure. Conformational preferences may provide insight into the targeting of the insertion mechanism to certain areas of the genome. Detailed characterization of the substrate interaction landscape of EN also informs ideal sites for small molecule inhibitors, for which our lab has established a variety of methods to identify and characterize. Here, we have completed a crystallographic fragment screen and utilized computational strategies to build upon promising hits. Going forward, further in vitro and in vivo experimentation with candidate inhibitors aims to not only identify compounds with greater potency than those currently available, but also to progress our understanding of EN-substrate interactions. Probing the mechanism of EN target specificity and its relation to retrotransposition activity or inhibition patterns is a promising route to sharpening our understanding of LINE1's life cycle and how to interfere with it effectively.

ASSEMBLY OF THE TN7 TARGETING COMPLEX AT SITES OF DNA REPLICATION

Shreya S Krishnan[1,3], Yao Shen[1,3], Nuwani W Weerasinghe[2,3], Rodolfo Ghirlando[4], Christopher Thibodeaux[2,3], Alba Guarné[1,3]

[1]Department of Biochemistry, McGill University, Montreal, Canada, [2]Department of Chemistry, McGill University, Montreal, Canada, [3]Centre de recherche en Biologie Structurale, McGill University, Montreal, Canada, [4]Laboratory of Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD

The bacterial Tn7 family of DNA transposons is known for its highly specific and programmable targe-site selection mechanisms. The canonical Tn7 element is directed into two specific target sites enabled by five element-encoded proteins: TnsABC define the core transposase machinery, while TnsD and TnsE are target-site selector proteins. TnsD recognizes a conserved sequence in the bacterial chromosome facilitating transposition into a safe haven with very high frequency. TnsE recognizes sites of DNA replication with a preference for 3'-recessed DNA and promotes transposition into conjugal plasmids. In this study, we used integrative structural biology approaches to elucidate how TnsE targets actively replicating DNA and promotes Tn7 insertions with a strict spacing and orientation. Using native mass spectrometry and analytical ultracentrifugation (AUC), we find that TnsE binds 3'-recessed DNA both as a monomer and a dimer, with gain-of-function TnsE variants promoting DNA-induced dimerization. Small-angle X-ray scattering (SAXS) and X-ray crystallography confirm that TnsE dimerizes on DNA, with the C-terminal domain of the protein binding to the duplex portion of the DNA substrates. This organization leaves the N-terminal domain of TnsE free to recruit TnsC to target sites – a process that we have visualized using cryo-electron microscopy. Collectively, our work allows us to propose a model for TnsE-mediated target-site selection that explains how the modular assembly of TnsE recruits TnsC to target sites and, in turn, directs Tn7 insertions to sites of active DNA replication.

UNVEILING THE UNIQUE EPIGENETIC REGULATION OF INDIVIDUAL L1S AT NUCLEOTIDE RESOLUTION.

Sophie Lanciano, Arpita Sarkar, Gael Cristofari

University Cote d'Azur, INSERM, CNRS, Institute for Research on Cancer and Aging of Nice (IRCAN), Nice, France

LINE-1 (L1) retrotransposons are abundant in the human genome and linked to disease and evolution. Normally silenced by DNA methylation, they are transcribed and eventually mobilized in many cancers. Given the repetitive nature and the dispersion of L1 sequences throughout the genome, deciphering the distinctive epigenetic regulation of individual copies has been challenging.
Recently, we developed bs-ATLAS-seq, a short-read targeted bisulfite sequencing strategy that can both locate and define the DNA methylation states of L1 elements genome-wide. This strategy unveiled L1 DNA methylation variation across cell types, families, and individual loci, elucidating key regulatory principles [1]. Notably, we observed that the methylation state of L1 elements can affect its proximal flanks, but only over a short distance. Additionally, L1 hypomethylation is accompanied by the binding of specific transcription factors, such as ESR1 (the estrogen receptor), in a breast cancer cell line, which drive the expression of L1 and chimeric transcripts. Moreover, we found that L1 hypomethylation alone is typically insufficient to trigger L1 expression due to redundant silencing pathways. However, studying regulations acting at the most internal part of the L1 promoter remained difficult due to short-read limitations.
To further explore the association between L1 chromatin compaction, DNA methylation, internal sequence variation, transcription factor binding, and expression, we have now combined Cas9-targeted nanopore sequencing and nanopore sequencing of nucleosome occupancy and methylome (nanoNOMe). This strategy allows us to obtain the entire sequence of L1 promoters genome-wide, their DNA methylation state at individual CpGs, as well as chromatin accessibility across L1 elements. By applying machine learning models, we inferred genomic features and transcription factors associated with L1 expression.
Despite L1 elements being extremely repetitive and each copy very similar to others, particularly for the youngest human-specific L1 elements, our results underline the importance of individual copy location, internal sequence, and regulation in understanding L1 reactivation in somatic cells.

[1] Lanciano S et al. Cell Genom. 2024 Feb 14;4(2):100498.

This work was supported by Agence Nationale de la Recherche (ANR-11-LABX-0028; ANR-15-IDEX-0001; ANR-19-CE12-0032; ANR-21-CE12-0001), Institut National Du Cancer (INCa PLBIO 2020-095), and other grants from the Canceropôle PACA, INCa and the Region Sud (Projet Emergence), Inserm (GOLD Cross-cutting Program on Genomic Variability), and CNRS (GDR 3546).

TRAFFIC JAM ACTIVATES THE *FLAMENCO* piRNA CLUSTER LOCUS AND THE PIWI PATHWAY TO ENSURE TRANSPOSON SILENCING AND *DROSOPHILA* FERTILITY.

Austin Rivera[1], Jou-Hsuan Roxie Lee[1], Shruti Gupta[1], Linda Yang[1], Raghuveera Goel[1,2], Joseph Zaia[1,2], Nelson C Lau[1,3]
[1]Boston University Chobanian and Avedisian School of Medicine, Biochemistry and Cell Biology, Boston, MA, [2]Boston University, Proteomics Service Center and Center for Network Systems Biology, Boston, MA, [3]Boston University Chobanian and Avedisian School of Medicine, Genome Science Institute, Boston, MA

Flamenco (Flam) is the most prominent piRNA cluster locus expressed in *Drosophila* ovarian follicle cells, and it is required for female fertility to silence *gypsy/mdg4* transposons. To determine how *Flam* is regulated, we used promoter-bashing assays in OSS cells to uncover novel shadow enhancer (SE) sequences within the first exons of *Flam*. We confirmed the SE relevance in vivo with new *Drosophila Flam* deletion mutants of these regions that compromised *Flam* piRNA expression and lowered female fertility from activated transposons. Our proteomic analysis of proteins associated with these SE sequences discovered the transcription factor *Traffic Jam* (*TJ*). *Tj* knockdowns in OSS cells caused a decrease in *Flam* transcripts, *Flam* piRNAs, and multiple Piwi pathway genes. A TJ ChIP-seq analysis from whole flies and OSS cells confirmed TJ binding exactly at the SE that was deleted in the new *Flam* mutant as well as TJ at multiple Piwi pathway gene enhancers. Surprisingly, TJ also bound the Long Terminal Repeats of transposons that had decreased expression after *Tj* knockdowns in OSS cells. Our study reveals the integral role TJ plays in the on-going arms race between selfish transposons and their suppression by the host Piwi pathway and the *Flam* piRNA cluster locus.

ELUCIDATING FACTORS INFLUENCING THE 5’ TRUNCATION OF L1 THAT OCCURS DURING INTEGRATION

SePil Lee, Abhishek Anand, Henry L Levin

Division of Molecular and Cellular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD

Long interspersed nuclear element-1 (LINE-1; L1) is a 6kb DNA sequence of an autonomous retrotransposon. L1 comprises 17% of the human genome with approximately 500,000 copies, representing the highest proportion of a transposable element in the genome. An Intact L1 DNA sequence contains regulatory regions, an internal promoter, and two coding sequences, ORF1 and ORF2. ORF1 protein forms trimers that bind L1 RNA and ORF2 encodes endonuclease and reverse transcriptase activities necessary for generating a nick that serves as the primer for L1 reverse transcription. The L1 sequences in the genome have a notable characteristic: most are severely truncated at the 5' end, resulting in a bimodal length distribution of full-length and 5' truncated elements. In truncated L1s, retrotransposition is inactivated due to the loss of promoter and coding sequences. The mechanism underlying 5’ truncation remains poorly understood.
Previous studies indicate L1 activity is influenced by multiple factors, including cellular factors and L1 sequences. Host pathways such as transcriptional gene silencing, innate immune responses, and DNA repair are known to repress L1 transposition. Additionally, studies of L1 reveal changes in transposition due to cis-DNA elements, including premature polyadenylation motifs in ORF2, splicing motifs, potential transcription factor binding sites, GC content, and amino acid sequences. However, the impact of these factors on 5’ truncation remains elusive.
To investigate the 5’ truncation mechanisms in retrotransposition, our laboratory developed a fluorescent-based reporter system that independently measures insertion of 5' and 3' L1 sequences. This system includes a mCherry reporter in the 5’ UTR and an EGFP in the 3’ UTR allowing for a more detailed examination of the factors affecting 5’ truncation.
Using this reporter system, we conducted knockdown experiments on 93 genes associated with DNA repair, innate immunity, DNA replication, cell cycle, and post-transcriptional gene silencing. MOV10 Knockdown consistently showed increased full-length integration. Additionally, we found 3 amino acid mutations affected 5’ truncation. Importantly, a mutation in the C-terminal domain of ORF1 increased the fraction of full-length insertion by 2-fold. To further study 5’ truncation, we are planning directed evolution experiments to identify L1 variants with increased levels of full-length integration. This study is expected to provide additional insights into the mechanisms of 5’ truncation.

EXPLORING THE GENETIC REGULATION AND INFLUENCE OF TRANSPOSABLE ELEMENTS IN VERTEBRATE HEALTH AND DISEASE

Aaron J Lemus[1,2], Alan Xu[1,3], Rajyk Bhala[1,2], Hengyu Zhou[1], Bryan Teefy[1], Bérénice A Benayoun[1,2,3,4]

[1]University of Southern California, Leonard Davis School of Gerontology, Los Angeles, CA, [2]USC Dornsife College of Letters, Arts, and Sciences, Molecular and Computational Biology, Los Angeles, CA, [3]USC Dornsife College of Letters, Arts, and Sciences, Quantitative & Computational Biology, Los Angeles, CA, [4]USC Keck School of Medicine, Biochemistry and Molecular Medicine, Los Angeles, CA

Transposable Elements (TEs) were historically dismissed as "junk DNA," due to their repetitive sequences, found in heterochromatin. Derepression of TEs has been seen in response to aging and in age-related disease (e.g. cancer and neurodegeneration). Accumulating evidence has revealed widespread derepression of TEs during aging across taxonomically distant model organisms. However, whether derepression of TEs is a byproduct of aging, or a driving factor of aging phenotypes, especially in vertebrate species, remains unknown. A key mechanism in TE repression is their regulation through heterochromatin formation. Importantly, the SUV39H1 gene encodes an enzyme that deposits trimethylation on lysine 9 of histone 3 (H3K9me3), a crucial repressive epigenetic modification of heterochromatin. Ubiquitous overexpression of the SUV39H1 homolog in flies was previously shown to lead to sustained TE transcriptional repression, and to extend health and lifespan. However, the relationship between H3K9me3-driven TE repression and healthy aging is conserved in vertebrates is unclear. The African turquoise killifish (Nothobranchius furzeri), a naturally short-lived vertebrate model (4-6months), provides an opportunity to investigate this link in a vertebrate model. Our preliminary data found that TE expression dramatically increases with aging across somatic tissues of both male and female turquoise killifish, making it an ideal model to investigate the impact of modulating TE expression on aging phenotypes. I hypothesize that progressive loss of TE repression, partially through loss of heterochromatin, contributes to the age-related functional decline in vertebrates. We propose to assess the causal role of TE expression on aging in the turquoise killifish model by manipulating TE control through genetic modulation of SUV39H1. Findings from this study will advance our understanding of TEs in aging and establish promising avenues for therapeutic interventions in age-related diseases.

USING MOBILE ELEMENTS AS GENETIC MARKERS FOR ANALYZING ADMIXTURE AMONG NEANDERTHAL, DENISOVANS, AND MODERN HUMANS

Sultan Mussakhan[1], Brooke Atamanyk[1], Grace Saad[1], Thien B Le[1,2], Ping Liang[1,2]

[1]Brock University, Department of Biological Sciences, St. Catharines, Canada, [2]Brock University, Centre for Biotechnologies, St. Catharines, Canada

The advent of next generation DNA sequencing technologies coupled with breakthroughs for working with ancient fossil DNA in last decades has made it reality for whole genome sequencing (WGS) tens of thousands of present-day humans from diverse ethnic groups, along with hundreds to thousands of ancient genomes for modern and archaic human groups including Neanderthal and Denisovans. Availability of such data has enabled us to start addressing long-standing questions regarding the relationship between archaic and modern humans. Standing out among the many surprises revealed by such work so far is the likely interbreeding events between Neanderthal and Denisovans with modern humans, which might have partially shaped the latter, detailing local adaptations and survivability for various present-day human populations. Studies in this field thus far have mainly utilized single nucleotide polymorphisms (SNP) as the genetic markers. Despite being the most abundant and easy to identify genetic markers, SNPs are identical-by-state (IBS) by nature, making them less ideal for analyzing certain intriguing and intricate questions, such as the fine details of the interbreeding and the later lineage-specific trajectories in modern humans. In this regard, transposable elements (TEs), aka mobile elements (MEs), are considered genetic markers superior to SNPs for being identical-by-descent (IBD) and abundant in number. However, archaic and ancient WGS data are mostly in a state of very short, highly mutated single reads because of the high level of DNA degradation and modification, limiting their use for analyzing ME polymorphism with existing tools. To overcome the challenges, we developed a new tool, called SingleME, tailored for working with archaic and ancient WGS datasets for identification of ME polymorphisms. With SingleME, along with other tools, we have so far analyzed 48 Neanderthal and 5 Denisovan genomes, along with ~750 representative modern human genomes from 1KGP, HGDP, and SGDP projects. Using 14 WGS datasets of the Mbuti group, known to be non-admixed with Neanderthals and Denisovans, as the baseline, we identified a total of ~1,550 MEs (~1,250 Alus, ~250 L1s, and ~40 SVAs) as the candidates of introgressed MEs between modern humans and Neanderthal/Denisovans with high confidence. At the meeting, we will present the detailed results from the analysis of these ME-markers, touching on questions including the number of interbreeding and their estimated timing, new insights on relationship of modern human populations, etc.

ROLE OF H3K9ME3 IN CONTROLLING TRANSPOSABLE ELEMENTS IN ADULT HAIR FOLLICLE STEM CELLS

Yu-Ching Liao, Cedric Feschotte, Tudorita Tumbar

Cornell University, Molecular Biology and Genetics, Ithaca, NY

Stem cells exhibit high potential to differentiate into diverse cell types, which is mirrored in their hyperplastic epigenome. Epigenetic regulation mediates stem cell potency without altering DNA sequences. Tri-methylation of histone H3 at lysine 9 (H3K9me3) promotes chromatin compaction, transcriptional repression, and transposable elements regulation (TEs) to safeguard genomic integrity. Most studies of TEs and H3K9me3 dynamics focused on early embryos or embryonic stem cells; however little is known about this dynamic in adult tissue stem cells, which are essential for tissue homeostasis and repair. Hair follicle stem cells (HFSCs) from the mouse skin is a powerful model to study adult tissue stem cells, as they undergo periodic activation to regenerate. Previous studies in the Tumbar laboratory have revealed a global decrease of H3K9me3 in HFSCs during stage transitions into quiescence in homeostasis. This result prompted us to investigate the status of TE activity at this stage. Specifically, we examined whether there was a decrease of H3K9me3 on TEs and whether it triggered TE activation. Analysis of ChIP-seq data for H3K9me3 suggests that despite widespread loss of H3K9me3 across the genome young and potentially active TEs retain H3K9me3 preferentially in HFSCs. The mechanisms underlying the selective maintenance of H3K9me3 on young TEs is unknown and warrants further investigation. This research aims to shed light on the intricate interplay between H3K9me3 and TE regulation in adult somatic stem cells during tissue homeostasis.

CONTRIBUTION OF AN SVA-DERIVED RNA TO X-LINKED DYSTONIA PARKINSONISM

Alan Lorenzetti*[1], Laura D'Ignazio*[1,2], Taylor A Evans*[1,2], Ellen B Penney*[3,4], Bareera Qamar[1], Arthur Feltrin[1], Alejandra E McCord[1], Andre Barbosa[1], Bonna J Sheehan[1], Bruno Araujo[1], Hunter H Giles[1,5], Yanhong Wang[1], Tomoyo Sawada[1], Michael Zabolocki[6,7], Cedric Bardy[6,7], D. Cristopher Bragg[3,4], Apua Paquola[1,2], Jennifer A Erwin[1,2,8]

[1]Lieber Institute for Brain Development, Division of Translational Neuroscience, Baltimore, MD, [2]Johns Hopkins University School of Medicine, Department of Neurology, Baltimore, MD, [3]Massachusetts General Hospital and Harvard Medical School, Department of Neurology, Boston, MA, [4]Massachusetts General Hospital, The Collaborative Center for X-linked Dystonia-Parkinsonism, Charlestown, MA, [5]Johns Hopkins University School of Medicine, Department of Genetic Medicine, Baltimore, MD, [6]SAHMRI, South Australian Health and Medical Research Institute, Adelaide, Australia, [7]College of Medicine and Public Health, Flinders Health Medical Research Institute, Adelaide, Australia, [8]Johns Hopkins University School of Medicine, Department of Neuroscience, Baltimore, MD

Up to 42% of the human genome is comprised of retrotransposons, a class of transposable elements (TEs) known to impact mammalian phenotypes. Among these, the non-autonomous SINE-VNTR-Alu (SVA) retrotransposons generates genetic diversity through insertion polymorphisms and length variation in short tandem repeat (STR) domains. This study focuses on the role of a TE in X-linked dystonia parkinsonism (XDP), a striatal neurodegenerative disease caused by an SVA antisense insertion within intron 32 of the *TAF1* gene. In this disease, the length of the XDP-causing SVA 5' STR (CCCTCT)n inversely correlates with the age of disease onset. By analyzing the phenotypic and transcriptomic profiles of XDP patient-derived and respective isogenic SVA-deleted striatal organoids, we found that the SVA insertion triggers key neurodegenerative features, such as transcriptional dysregulation, reduced neuronal activity, and apoptosis, whereas these abnormalities are mitigated in the SVA-deleted counterparts. Inspecting bulk RNA sequencing reads, we discovered an (AGAGGG)n repeat-containing RNA antisense to the XDP SVA that shows increased expression during organoid maturation and participates in R-loop formation. We characterized this transcript using rapid amplification of cDNA ends and determined the existence of 5' ends within the XDP SVA 5' STR region. We also observed supporting evidence for antisense 5' ends within SVA STRs by analyzing brain long-read RNA sequencing data from elderly donors, suggesting that this phenomenon can occur in vivo. Furthermore, we obtained preliminary evidence that the use of antisense oligonucleotides designed to knock down the (AGAGGG)n repeat-containing RNA can rescue the apoptotic phenotype in the XDP organoids. Our findings demonstrate that a tandem repeat-containing RNA derived from a retrotransposon can induce neurodegeneration and may be used as a target for prospective therapeutic development.

THE ROLE OF AGO2 AND AGO3 IN RNAi BALANCE AND EPIGENETIC INHERITANCE

Jason S Lynn[1,2], Evan Ernst[1,2], Samantha Blau[1], Harper Lowry[1], Jonathan Cahn[1,2], Jean-Sebastein Parent[1], Robert Martienssen[1,2]

[1]Cold Spring Harbor Laboratory, Plant Biology, Cold Spring Harbor, NY, [2]Howard Hughes Medical Institute, Cold Spring Harbor, NY

RNA interference (RNAi) is a highly conserved mechanism that regulates gene expression through Transcriptional Gene Silencing (TGS) and Post-Transcriptional Gene Silencing (PTGS). These two pathways exhibit profound differences in their persistence. Generally, PTGS represents a shorter-term silencing outcome, limited by the abundance of small RNAs and their cognate target mRNAs. In contrast, TGS, due to the amplification of small RNAs from the target locus after silencing, can be heritable within and across generations. The gene silencing outcome of RNAi in plants crucially relies on the interaction of small RNAs with Argonaute (AGO) proteins, forming AGO-effector complexes that target homologous RNAs. While seminal work has identified features associated with small RNA sorting into AGO, the overlap in binding preferences among closely related AGOs involved in TGS or PTGS blurs the lines for how plants control RNAi fidelity. Consequently, there is a growing need for a deeper understanding of the complex sorting mechanisms.

To address these questions, we leverage two model systems: *Arabidopsis thaliana*, with its extensive collection of epigenetic mutants and well-documented Argonaute literature, provides a valuable platform for investigating RNAi balance. The b1 paramutation system in maize offers unique insights into the role of AGO balance in transgenerational inheritance. Our results provide support for AGO2's involvement in suppressing ectopic TGS under environmental stress and in epigenetic inheritance.

TEMPERATURE-DEPENDENT MOBILITY OF TEs IN A HUMAN FUNGAL PATHOGEN: INVESTIGATING THE MECHANISMS OF REGULATION

Anna Mackey[1], Vesper Fraunfelter[1], Samantha Shaltz[1], Callan Schroeder[1], John Perfect[2], Sue Jinks-Robertson[1], Asiya Gusa[1]

[1]Duke University, Molecular Genetics and Microbiology, Durham, NC, [2]Duke University, Department of Medicine, Durham, NC

Transposable elements (TEs) have been shown to impact transcriptional networks, virulence traits, genome architecture, and evolution across eukaryotes. My work focuses on the impact of stressors and genetic background on TE mobilization dynamics in the human fungal pathogen *Cryptococcus neoformans*. *C. neoformans* is an environmental yeast which causes infections in immunocompromised individuals, accounting for up to 20% of annual AIDS-related deaths. We have screened a collection of serial clinical isolates from 15 patients with recurrent cryptococcal infections for TE activity during growth at 30°C and the stressful condition of mammalian body temperature (37°C) using a reporter gene approach. This screen has revealed interesting temperature and isolate-specific TE dynamics, including identification of a heat stress-activated LTR retrotransposon Tcn3 with putative heat shock binding elements in the LTR. Additionally, we identified clonally related isolates recovered from the same patient displaying a >50-fold difference in TE insertion rates into our reporter gene. To understand the mechanisms driving differential TE activity in these isolates we utilized long read sequencing and de novo genome assembly to assess copy number and genomic context of active TEs. Ongoing work is testing two hypotheses 1) heat-activated Tcn3 hijacks the host's heat shock response to induce expression and mobilization during growth at 37°C and 2) an amino acid substitution in a putative histone lysine methyltransferase is causing derepression of multiple TE families. This work will deepen our understanding of mechanisms governing stress-induced genome instability and TE regulation in a pathogenic non-model organism.

CARROT MITEs PROVIDE BINDING SITES FOR A CIRCADIAN CLOCK TRANSCRIPTION FACTOR, LHY.

Alicja Macko-Podgórni, Wojciech Wesołowski, Kinga Zygmuntowicz, Kornelia Kwolek, Marcelina Skrabucha, Dariusz Grzebelus

University of Agriculture in Krakow, Dept. of Plant Biology and Biotechnology, Faculty of Biotechnology and Horticulture, Krakow, Poland

The circadian clock in plants plays a pivotal role in orchestrating various growth and cellular processes. Among the key regulators of this clock is Late Elongated Hypocotyl (LHY) transcription factor, which not only suppresses the activity of afternoon and evening-phased genes, thus regulating the circadian clock itself, but also exerts influence on the expression of genes involved in growth, development, and regulatory pathways. Our recent research has revealed a significant enrichment of canonical binding sites of LHY within sequences of carrot Miniature Inverted-repeat Transposable Elements (MITEs).
In this study, we used DNA affinity purification sequencing (DAP-seq) to experimentally identify and characterize LHY binding sites across the genome. In addition, we sequenced the transcriptomes of carrot plant leaves exposed to four hours of heat, cold and salinity stress.

A total of 11,779 binding peaks were identified, of which 20% were located in promoter regions. Gene Ontology (GO) analysis of genes with LHY binding sites in promoter regions revealed enrichment of genes involved vegetative to reproductive phase transition of meristem, rhythmic processes, flower development, cell division, chromatin binding and signal transduction. Key KEGG pathways included circadian rhythm in plants and carotenoid biosynthesis, leading to ABA biosynthesis. Of the 2,346 genes with DAP-seq peaks in promoter regions, 74% were expressed, and of these, 694 genes showed altered expression in response to at least one abiotic stress. It is noteworthy that LHY expression significantly decreased under heat stress, while 248 of the 694 differentially expressed genes (36%), including TOC1 and other circadian clock-related genes that should be regulated by LHY, were up-regulated under heat stress.
We also compared the experimentally identified LHY binding sites with the locations of MITE transposons. We found that 1,428 (12%) of DAP-seq peaks overlapped with MITE copies, and 41% of these overlapped peaks were within elements of the Tourist_15 family, comprising 56% of all elements in this family. The proportion of Tourist_15 copies in promoter regions was even higher, constituting 48% of all MITEs in these regions. Thus, Tourist_15 MITEs can provide LHY binding sites, and if mobilized, they may rewire the circadian clock network and potentially accelerate adaptation to abiotic stresses.

CHARACTERIZING THE COOPERATIVE ASSEMBLY AND NEUROBIOLOGIC ROLE OF DOMESTICATED RETROELEMENT PNMA CAPSIDS

Victoria J Madigan[1,2,3,4], Feng Zhang[1,2,3,4]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, [2]Massachusetts Institute of Technology, Brain and Cognitive Sciences, Cambridge, MA, [3]Massachusetts Institute of Technology, Biological Engineering, Cambridge, MA, [4]Howard Hughes Medical Institute, Cambridge, MA

Domesticated retroelements are abundant in the human genome, some of which contain capsid domains with structural similarity to retroviral capsid proteins. Recent work has shown that domesticated retroelements expressed in the central nervous system (CNS) can assemble into virus-like particles capable of intercellular RNA transfer. While certain domesticated retroelements, such as Arc, have been shown to play potent roles in neurobiology, many remain poorly characterized. The paraneoplastic Ma antigen (PNMA) family of domesticated retroelements are highly expressed in the mammalian CNS, and eutherians have over a dozen PNMA genes with gag-like capsid domains. Some PNMA proteins assemble into capsid structures. Mouse and human PNMA2 have been shown to form non-enveloped, T=1 icosahedral capsids. Given predicted structural similarities between PNMA capsid proteins, we hypothesized that PNMA family members might co-assemble into chimeric capsid proteins. Using immunoprecipitation from mammalian cells, we found that PNMA2 interacts with PNMA7a and PNMA7b. We confirmed the interaction between PNMA2 with PNMA7a and PNMA7b *in vitro* with recombinantly produced proteins, and found that PNMA7a and PNMA7b can be incorporated into PNMA2 capsids. By investigating single-cell RNAseq datasets of the mammalian brain, we found that PNMA2, PNMA7a, and PNMA7b are highly co-expressed in neurons. We visualized the intracellular localization of PNMA2, PNMA7a, and PNMA7b with immunofluorescence of primary neurons, finding vesicle-like puncta of all three PNMAs in the neurites, cell body, and nucleus. To examine the function of these capsids in the mammalian brain, we are investigating their impact on neuronal gene expression, their binding partners, and their trafficking in the CNS. We have optimized CRISPRa of PNMA2, PNMA7a and PNMA7b in cultured neurons, and are currently dissecting their individual and collaborative impacts on neuronal gene transcription. Given the putative role of PNMA7a and PNMA7b as transcription factors, we are also performing an analogous Cut and Run assay to map the association of PNMA2, PNMA7a, and PNMA7b with the neuronal genome. Using affinity-purification mass spectrometry, we have identified binding partners of PNMA2 in the mouse CNS, finding a striking association of PNMA2 with synaptic vesicle machinery, SNARE membrane fusion machinery, and components of the DNA damage response (DDR). Our continued efforts aim to understand how these domesticated capsids influence neuronal homeostasis and how they utilize SNARE machinery for intercellular trafficking. Given the critical role of both vesicle trafficking and DDR in neuronal homeostasis and neurodegeneration, we believe that understanding how PNMA capsids interface with these processes will unveil fascinating new retroelement biology and potentially guide design of novel biomedical tools.

UNRAVELING THE MOLECULAR MECHANISMS OF *PIGGYBAT*-MEDIATED DNA TRANSPOSITION

Raju Mandal, Alison B Hickman, Laurie Lannes, Christina Hong, Fred Dyda

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD

DNA transposons are mobile genetic elements that can move from one location to another within the host genome. They play a significant role in genome evolution and are found in a wide range of organisms, from prokaryotes to eukaryotes. However, most DNA transposons in higher organisms, including humans, are inactivated due to mutations. The *piggyBat* DNA transposon is the only DNA transposon still active in mammals, which was identified in the little brown bat (*Myotis lucifugus*) and it belongs to the *piggyBac* superfamily. The *piggyBat* transposase system encodes a transposase that catalyzes its movement within the genome by recognizing its Left End and Right End terminal inverted repeats, liberating the element through double-strand breaks and inserting the transposon into a TTAA tetranucleotide at another genomic location. Studies have shown that the *piggyBat* transposase is active in yeast and human cells. Recently, the *piggyBat* transposase system has been used as a tool for genomic application, despite not knowing its in-depth transposition mechanisms. Therefore, our goal is to elucidate the molecular mechanisms behind *piggyBat*-mediated transposition using structural and cell biological data and to harness this knowledge to optimize the tranposase system for future genomic applications. Our data suggest that the *piggyBat* transposase strongly interacts with the Left End (LE) donor DNA relative to the Right End (RE) donor DNA and that the active assembly is a tetramer in the presence of oligonucleotides derived from *piggyBat's* terminal inverted repeats. Furthermore, our cell biological data indicate that transposition activity can be significantly increased by shortening both the LE and the RE and by modifying the *piggyBat* transposase.

ROLE OF HOST-TRANSPOSASE FUSION PROTEIN POGZ IN 3D GENOME ARCHITECTURE OF THE DEVELOPING BRAIN

Natasha Ann F Mariano, Cédric Feschotte, Eirene Markenscoff-Papadimitriou

Cornell University, Molecular Biology and Genetics, Ithaca, NY

Brain development is a complex process orchestrated by multiple factors, including gene regulation, expression, and chromatin organization. How the three-dimensional structure of the genome is organized and altered is essential for transcriptional regulation during neurogenesis. Yet, proteins that coordinate changes in the cis-regulatory architecture during mammalian neurodevelopment remain unexplored. Pogo Transposable Element Derived with ZNF Domain (POGZ) is a highly expressed transcriptional regulator during early development of several tissues, including the brain, and *de novo* mutations in *POGZ* have been directly implicated in neurodevelopmental disorders, including autism. At the molecular level, POGZ protein function remains poorly understood and likely multifaceted since it has been associated both with transcriptional activation and repression. Its multifaceted role likely reflects its chimeric domain architecture with its N-terminal region derived from a host zinc finger protein and its C-terminal region from the co-option of a complete transposase. This fusion event provides POGZ with two putative DNA-binding domains as well as an intact transposase catalytic core, making it mechanistically plausible for POGZ to bind DNA, multimerize, and form a synaptic complex with other proteins including itself. Thus, we hypothesize that POGZ is capable of looping chromatin to assist in establishing enhancer-promoter interactions during neurogenesis. To begin testing this hypothesis, we generated chromatin capture contact data (Micro-C) in the developing cortex of Pogz wildtype and knockout mice. Preliminary data show that *Pogz* knockout mice exhibit loss of topologically associated domains in a small subset of early embryonic cortex regions containing POGZ-dependent neuronal genes and high-frequency POGZ binding sites. These results suggest that POGZ plays a critical role in regulating the chromatin landscape during neurogenesis for proper cis-regulatory interactions during neurogenesis.

NUCLEIC ACID BINDING PREFERENCE OF A NUCLEOCAPSID-LIKE DOMAIN FROM THE DOMESTICATED RETROTRANSPOSON PEG10

Autumn M Matthews, Holly H Black, Alexandra M Whiteley

University of Colorado Boulder, Biochemistry, Boulder, CO

In recent years, there has been a growing appreciation for the role of retroelements in driving various neurodegenerative disorders. For example, the domesticated retrotransposon, Paternally Expressed Gene 10 (PEG10), derived from the Ty3/Gypsy family of retrotransposons, has recently been implicated in both Amyotrophic Lateral Sclerosis (ALS) and Angelman syndrome. Although the molecular mechanism remains unclear, our lab has identified a role for PEG10 in influencing neuronal gene expression. Like its retroviral ancestors, *PEG10* mRNA encodes for a *gag* and a *pol* region separated by a programmed -1 ribosomal frameshift, allowing for the generation of a gag or a gag-pol protein from the same mRNA. The gag region contains a capsid and zinc finger domain, while the pol region contains an aspartic protease domain; however, it lacks other enzymes necessary for replication. We found that upon proteolytic self-cleavage, PEG10 generates a nucleocapsid (NC)-like protein fragment that localizes to the nucleus and alters transcript levels of genes in neuronal pathways. However, the mechanism by which PEG10 NC regulates gene expression remains unknown. Detailed analysis of RNA-Seq data identified a putative role for PEG10 NC in specific alternative splicing events. While there were no global changes to particular classes of splicing events after overexpression of PEG10 NC, some genes exhibited alternative splicing. To investigate how PEG10 NC regulates gene expression and alternative splicing, we have begun to characterize its nucleic acid binding preference. A recent study from Ray et al. 2023 used a large-scale screen to identify a proposed UUUUGGG RNA motif to which PEG10 binds. In collaboration with the Ed Chuong lab at CU Boulder, we found over 650,000 instances of this motif scattered throughout the genome and confirmed this motif is present in the genes that were most upregulated by PEG10 NC overexpression. Using fluorescent electrophoretic mobility shift assays, we show that recombinant PEG10 NC preferentially binds to this UUUUGGG RNA motif and identified 13-mer RNA motifs within *DCLK1* and *TXNIP* to which PEG10 NC binds. We also show that while PEG10 NC binds to this 13-mer RNA motif within *DCLK1*, it does not bind to a single-stranded DNA probe of that same sequence, highlighting a preference for RNA. These studies suggest a role for PEG10 NC in nuclear regulation of RNA biology. Future studies will focus on the mechanism by which accumulated PEG10 NC regulates RNA processing, as well as how this affects neuron biology. Ultimately, this research provides insights into how this novel factor may contribute to ALS disease progression.

FUNCTION AND EVOLUTION OF HOST-TRANSPOSASE FUSION GENE *POGZ* INCLUDING ACQUISITION OF AN ADDITIONAL TRANSPOSASE DOMAIN IN FISH

John McCormick[1], Natasha Ann F. Mariano[1], Vinh Truong[2], Eirene Markenscoff-Papadimitriou[1], Cedric Feschotte[1]

[1]Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY, [2]St. Jude Children's Research Hospital, Department of Structural Biology, Memphis, TN

The cooption of DNA transposons has repeatedly given birth to essential host genes over the course of evolution resulting in a growing number of documented biological innovations. While the human genome contains dozens of transposase-derived genes including *RAG1*, which is responsible for V(D)J recombination, many lack well-defined functions and the evolutionary trajectories for domestication are incompletely understood. *POGZ* is a host-transposase fusion gene with host-derived tandem zinc fingers in the N-terminal portion and a pogo transposase-derived C-terminus. *Pogz* is required for mouse development and mutations are associated with neurodevelopmental disorders in humans, yet the function and evolution of *POGZ* are not well understood. The fusion event that led to the birth of *POGZ* can be traced back to the common ancestor of jawed vertebrates, ~450 Mya. We leveraged the genome sequence of 1273 vertebrate species to examine in detail the evolutionary history of *POGZ*. 1237/1273 (97%) of tetrapod POGZ orthologs maintain all domains of the ancestral transposase, including the HTH DNA binding domain, catalytic core, and an intact catalytic triad DDD, suggesting that POGZ may have retained some catalytic activities. To test this idea, we are assaying transposition activities of the human *POGZ* transposase domain in human cells. In addition, we identify intriguing features of *pogz* evolution across fish. First, we observe a change in the catalytic core DDD motif to DDE in the common ancestor of bony fish ~ 425 My ago with a subsequent whole genome duplication giving rise to *pogza* and *pogzb* in teleost fish. The fish DDE motif has reverted to DDD independently at least twice, once in *pogza* in the common ancestor of medaka and cichlids, and once in platyfish *pogzb*. Second, we find *pogz* acquired an additional piggyBac transposase domain in the common ancestor of all bony fish, which was lost in teleost *pogza*. The piggyBac-containing *pogzb* isoform is alternatively spliced and co-expressed with the *pogzb* isoform lacking the piggyBac exon in larval zebrafish. Third, both *pogza* and *pogzb* display signatures of rapid evolution in fish species, indicated by longer than expected branch lengths. Fourth, in the medaka lineage *pogza* has acquired four introns within the ancestrally intron-less pogo-derived region. Taken together these observations paint a picture of dynamic evolution of *pogz* homologs across fish, which may indicate roles for *pogz* in fish adaptation or an ongoing genetic conflict. These divergent evolutionary trajectories in fish may serve to inform *pogz* function and illuminate general principles of transposon domestication.

TESTING MODELS OF INSERTIONAL BIAS FOR A CENTROMERE-ENRICHED NON-LTR RETROELEMENT

Tyler McDermott[1], Bianca Planeta[1], Barbara Mellone[1,2]

[1]University of Connecticut, Molecular and Cell Biology, Storrs, CT,
[2]University of Connecticut, Institute for Systems Genomics, Storrs, CT

The centromere is an essential chromosomal locus that serves as the site of assembly for the kinetochore, which in turn attaches to spindle microtubules to mediate accurate chromosome segregation during cell division. Centromeres are composed of specialized chromatin containing the essential histone H3-variant CENP-A and highly repetitive DNA. Retroelements have been found associated with centromere across taxa, but how these elements accumulate at centromeric chromatin is unknown. The non-LTR retroelement *G2/Jockey-3* (henceforth *Jockey-3*) is the only genetic element shared among all five centromeres of *Drosophila melanogaster*. *Jockey-3* is the most highly enriched repeat in CENP-A chromatin immunoprecipitations and, although copies of it are found throughout the genome, it is significantly enriched at centromeres. *Jockey-3* is also present at the centromeres of sister *Drosophila* species, despite the high divergence in their centromeric satellites. To determine if *Jockey-3* preferentially targets the centromere for reinsertion, we designed transgenic fly lines containing an engineered copy of *Jockey-3* (*eJockey-3*) under the control of an inducible promoter, and whose presence can be tracked by SNP barcodes and epitope tagged ORFs. Upon activation of full-length *eJockey-3*, we observe lethality and infertility following ubiquitous and germline expression, respectively. Immunofluorescence analysis shows *Jockey-3* ORF1p and ORF2p co-localize with CENP-A at a significant rate compared to a control non-LTR retroelement, *R2*, which targets rDNA for transposition. Using DNA damage as a proxy for retroelement activity, we observed significantly more γH2Av signal at centromeres for *eJockey-3* compared to *R2*. To confirm actual transposition, we designed a digital droplet PCR (ddPCR™) assay to detect *eJockey-3*'s unique barcode and find copy number variation in induced tissues. Collectively, these findings are consistent with the hypothesis that *Jockey-3* preferentially targets the centromere for transposition. Additional experiments are underway to further support these data.

THE ROLE OF CELLULAR HOST FACTORS IN NON-LTR RETROTRANSPOSON GENE INSERTION.

Jeremy J McIntyre, Connor A Horton, Kathleen Collins

University of California, Berkeley, Molecular and Cell Biology, Berkeley, CA

Recently we developed a method called PRINT (Precise RNA-mediated INsertion of Transgenes) for safe and generalizable transgene supplementation of the human genome. It exploits a eukaryotic protein from the R2 family of non-LTR retrotransposons, which can copy a co-transfected template RNA into a specific safe-harbor target site. The transgene payload sequence is inserted directly into the genome via the endonuclease and reverse transcriptase activities of the R2 protein through a process called target-primed reverse transcription (TPRT). However, the complete mechanism of TPRT in PRINT, and in non-LTR retroelement gene insertion generally, is still mysterious, particularly the steps involving 5' junction formation between the target site and transgene sequence and second-strand synthesis. The post-TPRT steps are likely mediated in part or entirely by cellular host factors. Sequencing of the 5' junctions of PRINT-inserted transgenes suggests there are multiple pathways for junction formation, some of which rely on template RNA 5' homology to the upstream target site sequence to produce precise, seamless junctions by annealing of target site and cDNA. However, some insertions occur without requirement of homology and thus form unproductive, mutagenic 5' truncated insertions, a feature observed in physiological mobility of non-LTR retrotransposons. Using high throughput siRNA screening with a focus on the cellular host factors involved in DNA synthesis and repair, we discover a remarkable integration of non-LTR transposon-protein mediated gene insertion with known pathways of DNA repair, with some pathways required generally and others differentially depending on the presence or absence of target site homology within the template RNA sequence. Data from our screens and follow-up characterization allow us to begin to build complete, multi-branched models of non-LTR retrotransposon and PRINT gene-insertion mechanisms in human cells. In addition, we leverage the discovery of these mechanisms to manipulate the cellular host factor impact on transgene insertion to improve the efficiency and efficacy of PRINT, enhancing its utility as a therapeutic tool.

EVOLUTIONARY ORIGINS OF ARCHAEAL AND EUKARYOTIC RNA-GUIDED RNA MODIFICATION IN IS110 TRANSPOSONS

Chance Meers[1], Hugo Vaysset[2], Jean Cury[2], Aude Bernheim[2], Samuel H Sternberg[1]

[1]Columbia University, Department of Biochemistry and Molecular Biophysics, New York, NY, [2]Institut Pasteur, Molecular Diversity of Microbes Lab, Paris, France

Guide RNAs direct enzymatic activities to target nucleic acids via base-pairing interactions, offering more reprogramming flexibility than protein-only systems. This adaptability led to the independent evolution of RNA-guided processes, one prominent example being the prokaryotic CRISPR-Cas systems that use guide RNAs for adaptive immunity against phages. Notably, recent findings have shown that the ancestors of single-effector Cas nucleases, such as Cas9 and Cas12, are found within bacterial transposons and also use guide RNAs to direct targeted DNA double-strand breaks, illustrating the emergence of guide RNA pathways within transposons. Remarkably, the innovation of guide RNA systems extends beyond CRISPR and transposon-encoded nucleases. In eukaryotic and archaeal cells, guide RNAs are essential components of the snoRNA-guided RNA modification process, directed by C/D box snoRNAs, which guide site-specific modifications of rRNA, crucial for proper folding, stability, and function of rRNA in ribosome assembly and protein synthesis. Our detailed phylogenetic and structural analyses have uncovered that these RNA-guided mechanisms originated from ancient IS110-like transposases in bacteria. Furthermore, recent biochemical and genetic studies by the Hsu and Ataide labs demonstrated that IS110 elements harness guide RNAs for directing DNA recombination, illustrating an evolutionary tactic where guide RNAs initially used for DNA recombination have been repurposed for directing precise RNA modifications. These insights underscore the pivotal role of transposons in the genesis and evolutionary development of guide RNA mechanisms, highlighting transposons, yet again, as fundamental innovators in the evolution of new molecular functions.

EXONIZATION OF MOBILE ELEMENTS AS A KEY SOURCE OF CHIMERIC TRANSCRIPTS IN GLIOBLASTOMA

Rafael Luiz V Mercuri[1,2], Thiago Luiz A Miller[1], Filipe F dos Santos[1,3], Matheus de Lima[4], Aline R Pozzo[2,4], Pedro A F Galante[1]

[1]Hospital Sírio-Libanês, Molecular Oncology, São Paulo, Brazil, [2]University of São Paulo, Internunities Program in Bioinformatics, São Paulo, Brazil, [3]University of São Paulo, Multicenter Program in Biochemistry, São Paulo, Brazil, [4]University of Manitoba, Cancercare, Winnipeg, Canada

Glioblastoma (GBM) is the most aggressive form of primary brain tumor in adults, with a median survival of approximately 15 months. Unfortunately, the prognosis for GBM has remained poor for decades. The genetic and transcriptomic diversity of GBM makes treatment challenging, emphasizing the need for a deeper understanding of its molecular foundations. Recently, retrotransposons have been found to impact gene expression and the creation of chimeric transcripts. While they may be linked to various cancers, their exact role in GBM is still not well understood. In this work, using paired-end RNA-Seq data from The Cancer Genome Atlas (TCGA) and GBM (U251) cell line, we investigated the exonization of mobile elements and formation of GBM specific chimera transcripts. For retroelement references, we used RepeatMasker annotations. To detect, describe, and measure the expression of chimeric transcripts, we developed FREDY, a tool designed to identify exonization of retrotransposable elements in RNA-seq data. First, we discovered 1012 chimeric transcripts associated with exonization of intragenic (introns located) mobile elements. In details, Alu elements exonization was the most frequent (~45%) and happened in the 5' most regions of transcripts. As expected, majorly (83%) of these exonizations involve Alu located in the opposite transcription direction of their host gene. Conversely, LINE1 elements and retrocopies of protein coding genes were mostly exonized (46% and 56.7%, respectively) in the 3' most regions of host genes and in the same transcriptional direction as their host genes. Interestingly, we identified two chimera transcripts candidates affecting the tumor suppressor gene ROS1. Furthermore, when comparing potential protein domains, we found that the coding sequences of these chimeric transcripts deleted two of the ROS1 domains, suggesting a premature termination of transcription and potentially translation. Additionally, we quantify that these chimera transcripts account for 84% of the total expression of the ROS1 gene in the GBM samples which were identified. We have developed a framework to identify formation of chimeric transcript in cancer by the exonization mobile elements. We apply our strategy to GBM samples and cell lines and we have identified chimeric transcripts of tumor suppressor gene (ROS1) that may play a crucial role in GBM carcinogenesis.
Support: 2020/02413-4

sideRETRO: A PIPELINE FOR IDENTIFYING SOMATIC AND POLYMORPHIC INSERTIONS OF RETROCOPIES (retroCNVs)

Thiago L. A. Miller[1], Fernanda Orpinelli[1], J. Leonel L. Buzzo[2], Pedro A. F. Galante[1]

[1]Hospital Sírio-Libanês, Instituto de Ensino e Pesquisa, São Paulo, Brazil, [2]Fleury, Genômica, São Paulo, Brazil

The advent of Next-Generation DNA sequencing technologies, facilitated by the foundational work of the Human Genome Project, has revolutionized genetic research by enabling cost-effective, high-throughput sequencing data generation. These advancements, coupled with enhanced computational processing capabilities, have paved the way for large-scale genetic analyses, allowing the investigation of individual genomic variations that were previously inaccessible. Among these variations, retrocopies of protein coding genes — also known as processed pseudogenes — are particularly significant. These elements arise from the retrotransposition of a mature mRNA mediated by the LINE-1 enzymatic machinery and can exist as fixed elements, present in all individuals of a species, or as unfixed elements, which include polymorphic, germline, or somatic variations.

Despite their importance, the study of unfixed retrocopies (retroCNVs) has been hindered by a lack of specialized bioinformatics tools for their identification and annotation in NGS data. Addressing this gap, we developed sideRETRO, a computational tool designed to detect and annotate unfixed retrocopies in whole genome and exome NGS datasets. sideRETRO offers detailed annotations, including insertion coordinates (chromosome, insertion site, and DNA strand orientation), genomic context (exonic, intronic, or intergenic), genotyping status (present or absent), and haplotyping status (homozygous or heterozygous).

To validate its effectiveness, sideRETRO was rigorously tested on both simulated data and real sequencing data, the latter having been independently verified by an external research group. The results confirmed sideRETRO's robustness and accuracy in identifying and characterizing unfixed retrocopies, thereby filling a crucial methodological void in structural variant analysis.

In summary, this work introduces sideRETRO as a powerful and efficient tool for the comprehensive analysis of retrocopy insertions. By enhancing our ability to detect and study retroCNVs, sideRETRO contributes significantly to the field of genomics, enabling the generation of novel hypotheses and facilitating systematic investigations into structural genomic variations caused by unfixed retrocopies.

CONSEQUENCES OF L1 RETROTRANSPOSON EXPRESSION IN BREAST CANCER CELLS

Paolo Mita, Pooja Vikraman, Ze Chen, Manjunatha Kogenaru, Teresa Davoli
NYU Langone Health, Institute for systems genetics, New York, NY

Within the human genome, LINE-1/L1 retrotransposon is the only transposon able to autonomously copy and re-insert itself into a new genomic locus through the process of retrotransposition. L1 also provides the molecular machinery for the mobilization of SINEs such as Alus. Over 100 L1 copies are full length and are able to retrotranspose. These potentially dangerous copies of L1 are kept at bay by a plethora of cellular controls such as DNA methylation, epigenetic repression marks, APOBEC mutagenesis, translational control and RNA interference mechanisms. Despite the demonstrated mutagenic capability of L1 elements, its role in cancer has been often overlooked as a mere byproduct of DNA deregulation with tangential consequences for cancer establishment and development. However, advanced and cheaper sequencing techniques and novel insights in the cellular consequences of retrotransposons expression depict a more complex scenario. Reactivation of L1 creates nucleic acid products such as dsRNA, cytoplasmic dsDNA and DNA:RNA hybrids, able to trigger a viral response and interferon (IFN) activation. The consequences of these activation on the innate immune response and cancer microenvironment are being investigated and support the idea that L1 has a direct effect on cancer development.

We developed an in vitro model for L1 derepression in immortalized human Mammary Epithelial Cells (hMEC) that better recapitulate a physiological L1 re-activation. Characterization of these cells showed a L1-dependent increase of γH2AX and activated IFN-I pathway. Interferon (IFN)-related DNA damage resistant signature (IRDS) genes, a subgroup of interferon-stimulated genes (ISGs), previously found to promote resistance to DNA damaging chemotherapy and radiotherapy, are upregulated by L1 expression. Moreover, we found an increased transcription of genes associated with epithelial-mesenchymal transition (EMT) upon L1 expression. Scratch tests show that L1 expressing hMECs are indeed able to migrate faster across a scratch-induced gap in vitro compared to control cells. Despite the increased γH2AX signal and higher gap-migration across a scratch, L1 expressing hMECs display same doubling time and cell growth of control cells. Moreover, through high-density screens we are investigating how derepressed L1 can support or restrain cancer-relevant genes such as tumor suppressors and oncogenes.

We propose that L1 expression may functionally interact with the cellular and genetic context, steering tumor development towards more aggressive and therapy resistant cancers. We believe that a deeper understanding of the effects of L1 retrotransposon derepression could be harnessed for improved patient stratification, more effective diagnostics and the development of more comprehensive cancer therapies.

MECHANISMS OF TRANSCRIPTIONAL REGULATION AND CO-OPTION OF RETROELEMENTS AND ENDOGENOUS VIRUSES

Nikos Nikolopoulos, Kai Huang, Haoming Zhai, Daniela Gjorgjevikj, Yorgo Modis

University of Cambridge, Department of Medicine, MRC-LMB Molecular Immunity Unit, Cambridge, United Kingdom

The Human Silencing Hub (HUSH) guards the genome from pathogenic effects of retroelement expression. Composed of MPP8, TASOR and Periphilin, HUSH recognizes actively transcribed retroelements by the presence of long (>1.5-kb) nascent transcripts without introns. HUSH recruits effectors, including MORC2 and SETDB1, which remodel chromatin and deposit transcriptionally repressive epigenetic marks.
Here, we report crystal structures of the C-terminal domain (CTD) of MPP8 necessary for HUSH activity and the DomII-PIN region of TASOR. The MPP8 CTD consists of ankyrin repeats followed by a domain with structural homology to Siz/PIAS-family SUMO E3 ligases (1). AlphaFold3 modeling predicts that the MPP8 CTD forms extended interaction interfaces with a SPOC domain and a domain with a novel fold in TASOR. A cell-based HUSH reporter assay was used to validate the structural model of the MPP8-TASOR complex. The TASOR DomII-PIN structure reveals a homodimeric assembly, with implications for the overall oligomeric state of the HUSH complex. DomII has structural similarity to BRCT domains also found in proteins from the PARP family. Based on this similarity and a report that MORC2 binds PARP1 BRCT, we propose a role for TASOR DomII in the recruitment of MORC2. Together, our structural and biochemical data identifies novel structural elements in MPP8 and TASOR required for HUSH complex assembly and silencing, thereby fulfilling vital functions in controlling retrotransposons.
Some retroelements escape repression and are under purifying selection, suggesting they have acquired a useful cellular function. We have identified, Atlas virus, an endogenous nematode virus with an intact genome belonging to an ancestral retrovirus family. Our cryo-EM data show that the Atlas virus envelope glycoprotein has a class II viral membrane fusion protein fold not previously seen in retroviruses (2). We also find that Atlas Gag self-assembles into large icosahedral capsids. Atlas Gag has the same fold as retrovirus capsids but forms an icosahedral assembly distinct from those of other retroviruses (unpublished). We conclude that Atlas virus, with its bioactive proteins, has the potential to form viable virions or fulfill cellular functions.

1. Nikolopoulos N, Oda S, Prigozhin DM, Modis Y (2024) Structure and methyl-lysine binding selectivity of the HUSH complex subunit MPP8. bioRxiv DOI:10.1101/2023.12.21.572340
2. Merchant M, Mata CP, Liu Y, Zhai H, Protasio AV, Modis Y (2022) A bioactive phlebovirus-like envelope protein in a hookworm endogenous virus. Sci. Adv. 8:eabj6894

REWIRED REGULATORY PATHWAYS INVOLVING RETROTRANSPOSONS IMPACT TRANSLATION AND HOMEOSTASIS DURING REPRODUCTION AND PREIMPLANTATION DEVELOPMENT

Andrew Modzelewski[1], Chantal Diallo[1], Aparna Anand[2], Ting Wang[2], Lin He[3]

[1]University of Pennsylvania, Biomedical Sciences, Philadelphia, PA, [2]Washington University, Genetics, St. Louis, MO, [3]UC Berkeley, MCB, Berkeley, CA

Throughout evolution, viruses and mobile elements have hijacked cellular systems to copy/paste themselves throughout mammalian genomes, giving rise to Retrotransposons. While 2% of the human genome encodes proteins, ~48% is so-called “Junk DNA” that originates from these ancient insertions. Rare cases where retrotransposons disturb important gene functions has led to an “US versus THEM” model. However, this model may be an oversimplification, failing to reflect symbiotic relationships developed through co-evolution. Most retrotransposons are inactive through constant epigenetic surveillance or degraded through mutation, but a subset retains partial function where these relics can influence nearby gene expression and function by acting as enhancers, repressors, lncRNAs, and in some cases, fusion with a nearby protein-coding gene called “Chimeric Transcripts” that completely change the original function. Disruption of retrotransposon family expression results in embryonic lethality, suggesting unknown but essential functions.

The "ModzLab" studies the phenomenon of **Retrotransposon Reactivation** that is essential and intentional in early embryonic development. While silenced in most tissues, unintentional retrotransposon reactivation is observed in cases of epigenetic breakdown associated with aging, disease, and cancer. Re-analysis of single cell pre-implantation from 9 species revealed striking levels of dynamically expressed retrotransposon families with conserved reactivation patterns spanning all species investigated. Novel bioinformatics tools are used to map these extremely repetitive elements and using a highly efficient electroporation-based embryo editing method we call CRISPR-EZ, we generated 7 retrotransposon deletions mouse lines, each impacting different aspects of development and reproduction. Analysis of one of these deletions shows that genetic ablation of the *MT2B1* promoter driving *Rpl41* results in delayed global translation, causing stress-induced arrest and embryonic metabolism defects. This work aims to elucidate the role of $Rpl41^{MT2B1}$ in global translation as well as to gain insight into the oxidative and metabolic needs of the developing embryo. Once regarded as parasitic DNA, we are now realizing retrotransposons can be co-opted by the host as transient but essential gene regulatory elements in development, reproduction, and disease.

CONFORMATIONAL LANDSCAPE OF THE TYPE V-K CRISPR-ASSOCIATED TRANSPOSON INTEGRATION ASSEMBLY

Francisco Tenjo-Castaño, Nicholas Sofos, Luisa Stutzke, Piero Temperini, Anders Fuglsang, Tillmann Pape, Pablo Mesa, Guillermo Montoya

Structural Molecular Biology Group, NNF-CPR, University of Copenhagen, Copenhagen, Denmark

CRISPR-associated transposons (CASTs) are mobile genetic elements that co-opt CRISPR-Cas systems for RNA-guided DNA transposition. CASTs integrate large DNA cargos into the attachment (att) site independently of homology-directed repair, thus holding promise for eukaryotic genome engineering(1-2). Here, we present the 2.3 Å cryo-EM structure of the reconstituted 1MDa post-transposition complex of the type V-K CAST, containing Cas12k, TnsC, TnsB, TniQ, and S15, together with different assembly intermediates and diverse TnsC filament lengths, enabling the recapitulation of the integration complex formation(3). We show the existence of two checkpoints in the association of the Cas12k-transposon recruitment complex with the TnsC filament. In the first checkpoint, ZnFn1, ZnFn2, and the N-terminal tail of TniQ guide the interaction of the recognition complex by testing the association of TniQ with the sgRNA and the initial two molecules of the TnsC filament. The second checkpoint senses the proper orientation and alignment of the recruitment complex and filament by inserting the stem-loop of the sgRNA between the hairpin loops of the 4th and 5th TnsC protomers. The high resolution in the ATP pockets within the TnsC filament show why basal ATPase activity occurs due to hydrolysis-ready catalytic sites. The long C-terminal flexible linker anchors the TnsB subunits to the basic pockets in the outer face of the cylindrical TnsC filament through an acidic motif termed “the hook”. This association guides the correct orientation of the TnsB synaptic complex. Furthermore, TnsC filament depolymerization is enhanced by the complex of TnsB with the LTR(6)-SR(1) sites, mimicking the pre-integration complex, which promotes a larger increase leading to greater depolymerization not observed in the presence of the TnsB-STC complex. In summary, the filament can elongate but also can depolymerize in different degrees in the presence of various transposase complexes. The assemblies with different numbers of TnsC protomers explain the minor integration sites observed in the activity assays. The results of mutagenesis experiments probing the roles of specific residues and TnsB binding sites, show that transposition activity can be enhanced, and show that the distance between the PAM and att sites is determined by the interplay between the lengths of the C-terminal linker between the hook and TnsB and the TnsC filament. This singular model of RNA-guided transposition provides a foundation for repurposing CAST for genome-editing applications.

1. Faure, G., et al., (2019) Nat Rev Microbiol 17, 513-525.
2. Peters, J.E., et al., (2017) PNAS 114, E7358--E7366.
3. Tenjo-Castaño et al., (2024) Molecular Cell 84(12):2353-2367

MOLECULAR DOMESTICATION OF GYPSY-TY3 RETROTRANSPOSONS IN MOSQUITOES MAY DRIVE CRISPR-LIKE ADAPTIVE IMMUNITY INNOVATION

Daniel A Morales[1,2,3], Annette Montalvan[1,2], Matthew Goldman[1,2], Victoria K Eggers[1,2], Andre L Costa-da-Silva[1,2], Anthony J Bellantuono[1,2], Patrick T Dolan[3], Jason Pienaar[2], Janna L Fierst[1,2], Matthew DeGennaro[1,2]

[1]Florida International University, Biomolecular Sciences Institute, Miami, FL, [2]Florida International University, Department of Biological Sciences, Miami, FL, [3]National Institute of Allergy and Infectious Diseases - NIAID/NIH, Laboratory of Viral Diseases, Bethesda, MD

A long-held viewpoint of vector biologists has been that viral diseases have a negligible effect on the mosquito; however, research continues to uncover the significant fitness effects of arboviruses on mosquitoes. Although adaptive immunity in mosquitoes has been largely discounted in invertebrates, recent findings suggest that *Aedes* mosquitoes domesticated a subgenus lineage of Gypsy-Ty3 LTR-retrotransposons to integrate virus-derived sequence information, known as endogenous viral elements (EVEs), into CRISPR-like loci within their genomes. Small RNAs, known as piRNAs, produced from these loci have been shown to suppress viral replication, acting as effectors of an adaptive and heritable immunity.

To assess evolutionary histories in Culicomorpha, we prepared a deep phylogeny of several Culicomorphan genome assemblies, spanning >175 million years of divergence. Using the Ape package, we reconstructed the ancestral states of various host traits, including genome size and Piwi protein family expansion. The transposable element landscapes were annotated using the Extensive *de novo* TE Annotator (EDTA) and LTRs larger than 5kbp were searched for EVEs using an EVE-BLAST strategy. LTRs were aligned using MAFFT L-INS-i and phylogenetically reconstructed using IQTree2 with 1000 ultrafast bootstraps and 1000 SH-aLRT replicates. Phylogenetic clustering of EVE-positive LTRs were analyzed using the ETE3 Toolkits library. To further interrogate the molecular domestication of Ty3 LTRs in the Aedini tribe of mosquitoes, we performed an ancestral sequence reconstruction (ASR_{seq}) of the TE-suppressing Piwi protein family orthologs across Culicomorpha. Candidate ASR_{seq} predicted protein structures were analyzed for structural similarity using the root mean square deviation method.

Our results showed that EVE biogenesis via LTRs is restricted to a subgenus clade of the Ty3. We believe a whole-transposon domestication of EVE-capturing Ty3 subfamilies drove an overall advantageous expansion of the impressively large *Aedes* genomes (~1950Mbp) in response to viral attack, compared to its sister clade *Culex* which shows a conservation of a mid-sized genome (~500Mbp). This domestication likely resulted in within-host conflicts driven by complex trade-off dynamics between host genome size expansion, viral infectivity/pathogenicity, and LTR replicative capacity. Understanding the evolutionary history of Culicidae genome size expansion and the trade-off dynamics between TEs, viral elements, and the mosquito host is critical to illuminating how this alternative adaptive immunity has evolved.

ZNF267 IS INDUCED EARLY DURING INFLAMMATION RESPONSE AND REGULATES IT BY BINDING THE1C ELEMENTS

Athina N Triantou*, Santiago Morell*, Samuel J Daffern*, Juliette C Davis, Denali A Archer, Michael Imbeault

University of Cambridge, Department of Genetics, Cambridge, United Kingdom

KRAB Zinc finger proteins (KZFPs) is the largest subfamily of DNA binding proteins in humans. Each KZFP contains a KRAB domain that recruits TRIM28, which triggers the formation of heterochromatin at the target genomic region, and a zinc finger array of variable length which is used to locate and bind specific DNA sequences, mostly derived from transposable elements (TEs). Interestingly, many KZFPs are evolutionarily conserved and bind equally conserved remnants of TEs several millions of years after these transposable elements have lost their transposition potential. We hypothesize that these are domesticated regulatory platforms whose accessibility is epigenetically controlled by KZFPs in cell-types in which they are present, acting as epigenetic switches of these regions they bind. Analysis of publicly available large-scale transcriptional datasets show that different sets of KZFPs are indeed being differentially expressed in diverse cellular contexts.
Here we will show results of an analysis showing that a small subset of KZFPs is regulated during the immune response. We will discuss the dynamics of their co-evolution with their TE targets, the transcription factor binding sites found enriched at these loci, variations in epigenetic profiles between cell types, and the implications these results have for the potential of species-specific regulation of the immune system. We will focus on results obtained by a functional analysis of ZNF267, a primate-specific KZFP that is induced early during the inflammation response. It targets LTR elements of the THE1C family which contain transcription factor binding sites relevant for inflammation. Using a CRISPR KO system coupled with transcriptomic and epigenetic analysis, we show that ZNF267 plays an important role in controlling the response arising from these elements following an inflammation stimulus in macrophages, altering expression levels of nearby genes.

SINGLE-MOLECULE METHODS FOR DETECTING SOMATIC TE MOBILIZATION IN PLANTS

Andrea Movilli, Fernando Rabanal, Detlef Weigel

Max Planck Institute for Biology Tuebingen, Molecular Biology, Tuebingen, Germany

DNA methylation is a hallmark of TE repression in flowering plants. *Arabidopsis thaliana* mutants of *METHYLTRANSFERASE1* (*MET1*, *Dnmt1* homolog) are defective in DNA methylation maintenance, causing genome-wide hypomethylation and TE derepression. However, in germline studies of *met1* mutants, derepressed TEs are rarely found to increase their copy number. Whether this reflects selection against new insertions or differences in mobilizability is often unclear.

To unbiasedly assess mobilome activity, we measured somatic TE insertions/deletions – which should be subject to less selection than germline events – in 10 first-generation *met1* individuals in a non-reference accession background by long-read sequencing.

We found extensive variation in interindividual TE mobilization. At the same time, different TE families show heterogeneous levels of mobility across different individuals. As reported in previous analyses of germline transpositions in Col-0 (*A.thaliana* reference accession), ATCOPIA93 and VANDAL21 are among the most active families, nonetheless, we also detected annotated and unannotated TEs that were not previously known to mobilize.

Due to the plastic nature of the germline in plants, rare transposition events that take place in the soma have a non-zero chance to be passed on to further generations. PCR-free long reads uncovered extensive somatic TE mobilization, proving to be a high-resolution and unbiased method to characterize the activity of the mobilome.

GENETIC EXCHANGE NETWORKS BRIDGE DNA VEHICLES IN A BACTERIAL PATHOGEN

Héloïse Muller, Odion Ikhimiukora, Cheryl P Andam

University at Albany, State University of New York, Department of Biological Sciences, Albany, NY

Bacterial cells contain multiple types of DNA vehicles – one or few self-replicating chromosomes and a myriad of mobile genetic elements (MGEs) such as phage, plasmid, and transposon. MGEs are essential for mediating horizontal gene transfer between cells, thus shaping the genetic diversity and phenotypic innovations in bacteria. Each DNA vehicle exhibits a unique evolutionary history, even if they are all present within the same cell. However, little is known of the interactions between different types of mobile DNA vehicles because MGEs are often studied independently of each other. Here, we look at genetic sharing from the point of view of the DNA vehicle rather than the cell. We analyzed 936 genome sequences of clinical *Listeria monocytogenes* (*Lm*) sampled across 58 counties in New York, USA from 2000 – 2021. *Lm* is an opportunistic foodborne pathogen responsible for gastrointestinal illnesses and life-threatening invasive infections. We built genetic exchange networks based on sequence similarities and shared gene content between MGEs. Within the largest network are communities of strongly interconnected MGEs. Although most genetic sharing involve the same vehicle type (phage-phage, plasmid-plasmid, transposon-transposon), subsets within the network link different MGE types (plasmid-transposon, phage-plasmid, dsDNA phage-ssDNA phage). Phages and transposons did not share any genetic connections, suggesting impermeability between them. Genes involved in stress response are overrepresented in plasmids and transposons, which are often exchanged between the two DNA vehicles. Frequent but transient genetic exchanges between MGEs therefore shape inter-strain variation, adaptive potential, and population structure of bacteria at short timescales.

SOCIAL ISOLATION TRIGGERS ERV EXPRESSION AND NEURODEGENERATION IN A *DROSOPHILA* TDP-43 MODEL

Swetha MurthyGowda[1], Joshua Dubnau[1,2]

[1]Stony Brook Medicine, Department of Anesthesiology, Stony Brook, NY, [2]Stony Brook University, Department of Neurobiology and Behaviour, Stony Brook, NY

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are two fatal neurodegenerative disorders with motor defects and cognitive dysfunctions. Pathophysiological hallmarks of the disease include loss of nuclear localization and abnormal aggregation of TAR DNA-binding protein 43 (TDP-43), observed in neurons and glia in the affected brain regions of ~95% of ALS and ~40% cases of FTLD patients. TDP43 pathology also is associated with expression of retrotransposons and endogenous retroviruses (ERV), and accumulation of DNA damage. These phenotypes are observed in *Drosophila* models, and in post-mortem tissues of human subjects. We previously reported that TDP-43 aggregation triggered by over-expression in *Drosophila* glial cells types triggers spread of toxicity to neurons, leading to their death. Such spreading toxicity is a hallmark of disease, and has been attributed to a prion-like mechanism. But our prior work demonstrates that the intercellular spread also requires ERV expression. ERV expression and TDP-43 pathology are mutually reinforcing, providing a feedback to amplify toxicity. While the nature of the upstream triggers that set this process in motion are not understood, epidemiological evidence suggests that psychological stressors such as PTSD, anxiety, depression and loneliness may increase the risks of disease onset. We studied the effects of a chronic social isolation paradigm that was previously shown to cause sleep loss in flies. We find that early life chronic social isolation is sufficient to sensitize animals to subsequent induction of pathological levels of TDP-43 in glia. The social isolation acts as a primer, leading to more rapid induction of the MDG4-ERV, more rapid propagation of TDP-43 pathology to nearby neurons, and to shorter lifespan. Our findings have implications for the the association between loneliness and risk of neurodegenerative diseases in humans.

INVESTIGATING L1 ORF1p-BASED mRNA VACCINE IN A MOUSE MELANOMA model

Raj Nandani, Wenfeng An

South Dakota State University, Department of Pharmaceutical Science, Brookings, SD

Long interspersed element type 1 (L1) is a class of transposable elements that make up 17% of the human genome. Somatic L1 insertions promote genomic instability and have been implicated in the evolution of various human cancer types. A full-length L1 has two open reading frames (ORF1 and ORF2) that encode proteins essential for its replication in the genome. It has been well documented that L1 ORF1 protein (ORF1p) is overexpressed in tumor tissues as compared to the non-tumor tissues. ORF1p has also been associated with modulation of tumor microenvironment and activation of oncogenic pathways. Here, we are interested in targeting ORF1p as a tumor-associated antigen using an mRNA lipid nanoparticle (LNP) vaccine. We have successfully synthesized ORF1 mRNA and tested its integrity on agarose gels. Using a nanofabrication device, we have formulated LNPs with a particle size < 200 nm. The ORF1p mRNA is encapsulated with 80% efficiency and remains intact when examined after extraction from the LNPs. Transfection of mRNA LNP demonstrates functional translation into L1 ORF1p in cells. To investigate the effect of the ORF1p mRNA vaccine in vivo, we have established a stable mouse melanoma cell line overexpressing ORF1p. Ongoing studies focus on the safety and efficacy of the ORF1p mRNA LNP vaccine in a syngeneic mouse tumor model. Our work is a proof-of-principle study on testing the immunogenicity of L1 ORF1p. It has the potential to open the door for future development of ORF1p-targeting cancer immunotherapies.

THE INTERFERON-STIMULATED GENE PRODUCT HERC5 INHIBITS HUMAN LINE-1 RETROTRANSPOSITION

Kei Nishimori[1,2], Ahmad Luqman-Fatah[1], Yuzo Watanabe[3], Fuyuki Ishikawa[2], Tomoichiro Miyoshi[1,2]

[1]Laboratory for Retrotransposon Dynamics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [2]Department of Gene Mechanisms, Graduate School of Biostudies, Kyoto University, Kyoto, Japan, [3]Proteomics Facility, Graduate School of Biostudies, Kyoto University, Kyoto, Japan

Long INterspersed Element-1 (LINE-1 or L1) is the only active and autonomous retrotransposon in the human genome. A full-length and retrotransposition-competent L1 family can amplify its own sequences, which are inserted into other genomic loci by a “copy-and-paste” mechanism termed retrotransposition. Since L1 retrotransposition poses a threat to genome integrity and, occasionally, causes several diseases such as cancer and hemophilia, host cells have been evolving a defense system to inhibit L1 activity; however, its mechanisms are still poorly understood. The retrotransposition-competent L1 encodes two proteins, L1 ORF1p and ORF2p. L1 ORF1p is an RNA-binding protein with nucleic acid chaperone activity, while ORF2p has endonuclease and reverse transcriptase activities, both of which are required for retrotransposition. L1 ORF1p and ORF2p bind to L1 RNA in cis to form RiboNucleoProtein particles (RNPs), essential for L1 retrotransposition. To identify the host factors that restrict L1 retrotransposition, we performed a mass spectrometry analysis of the L1 ORF1p complex. This analysis revealed that L1 RNPs associate with multiple interferon-stimulated gene (ISG) products including HECT and RLD domain containing E3 ubiquitin protein ligase 5 (HERC5). HERC5 is an E3 protein ligase for ISGylation and restricts virus proliferation, both dependent and independent of ISGylation. Our findings demonstrate that HERC5 inhibits L1 retrotransposition through its RLD domain but not the HECT domain essential for ISGylation. This suggests an ISGylation-independent mechanism by which HERC5 inhibits L1. We will discuss the ISGylation-independent function of HERC5 and the biological significance of L1 inhibition by the innate immune system.

Keywords: Retrotransposon, LINE-1, ISGylation, innate immunity

EVOLUTIONARY DYNAMICS OF POLINTONS; A RARELY INVESTIGATED SUPERFAMILY OF TRANSPOSONS

Shumaila Noreen[1], Muhammad Umair Khan[2], Faisal Nouroz[2]

[1]Hazara University, Zoology, Mansehra, Pakistan, [2]Hazara University, Bioinformatics, Mansehra, Pakistan, [3]Hazara University, Bioinformatics, Mansehra, Pakistan

Based on the proliferation mechanisms of transposable elements, they are classified as Class I or retrotransposons and Class II or DNA transposons. The Class I elements are larger in sized in comparison to DNA transposon. Among DNA transposons, few supoerfamilies are frequently proliferating in their hst genomes, while few superfamilies are rarely available. Polintons, also referred as Mavericks are rarely investigated DNA transposons. Although discovered recently, these elements have now been found in several eukaryotic genomes because of the advancements & expansions in genome sequencing projects. The current research was conducted to investigate the evolutionary genomics and distribution of Polintons in various genomes. Around 102 Polintons from various species were collected from Repbase – GIRI database. The phylogenetic relationship and evolutionary history of the Polintons superfamily was investigated by using bioinformatics softwares. We created phylogenetic trees based on two major protein coding domains, as RVE integrase and DNA POL B. In both cases, the sequences were clustered in 2 clades, several sub-clades and groups with variable number of elements in each group. The cladograms represented the extensive genetic diversity and evolutionary history of these elements and allowed us to observe the intricate branching patterns and relationships within the Polinton superfamily. The cladogram provides a roadmap for exploring the relationships and dynamics of the Polinton superfamily in greater detail. This study enhances our understanding of evolutionary dynamics of Polintons in different organisms.
Keywords: Polintons, Transposons, RVE integrase, Phylogeny, Evolution.

THE GENOMIC BLUEPRINT OF TRANSPOSABLE ELEMENTS IN PLANTS

Faisal Nouroz[1,2], Shumaila Noreen[2,3]

[1]Hazara University, Bioinformatics, Mansehra, Pakistan, [2]University of Leicester, Genetics & Genome Biology, Leicester, United Kingdom, [3]Hazara University, Zoology, Mansehra, Pakistan

The transposable elements (TEs) or transposons are actively proliferating in plant genomes contributing to genome size duplication, evolution, diversification, and plasticity of their host genome. They are further classified into two classes; Class I or Retrotransposons and Class II or DNA transposons based on their transposition mechanism. Being highly diverse genomic sequences they are further divided into superfamilies, families and sub-families. The current work focused on the identification of diverse TEs in selected important plants by exploiting the bioinformatics, molecular and cytogenetic techniques. Several Bioinformatics tools/ softwares were exploited for the identification of these TEs. Of the LTR retrotransposons, the superfamilies like Copia, Gypsy, and retroviruses (caulimoviruses and pararetroviruses) were actively and abundantly proliferating in these plant genomes. The Copia and Gypsy superfamilies were found in high copy numbers followed by retroviruses (caulimoviruses and pararetroviruses) in Brassica, Musa, Citrus, and Datepalm. Among Non-LTR retrotransposons, the members from superfamilies SINES and LINES predominated, while DNA transposons were abundantly represented by CACTA, hAT, Harbinger, Mariner, Mutator and MITEs superfamilies, while Helitron, Politron and Maverik were rarely identified. PCR amplification of reverse transcriptase (RT) of Retrotransposons and transposase (TNP) of DNA transposons revealed their distribution among various genomes, with some elements were found to be specie or genera specific, while others were mobilized by horizontal transfer. The evolutionary relationship of these elements resolved them into superfamily and family specific lineages. The results enabled the characterization, annotation, evolutionary dynamics, and structural features of MGEs and their derivatives in plant genomes. These TEs can be utilized as best genetic markers to study the genetic diversity and evolutionary relationship of the plants.
Key words: Genome, Bioinformatics, Retrotransposons, Gypsy, Retroviruses, Evolution.

CONTRASTING EFFECTS OF HISTONE H2A VARIANTS ON THE ESTABLISHMENT OF TRANSPOSON-SPECIFIC SILENT MODIFICATION IN *ARABIDOPSIS*

Shoko Oda[1,2], Sayaka Tominaga[2], Shumpei Takeuchi[1], Tetsuji Kakutani[1], Taiko K To[1,2]

[1]The University of Tokyo, Biological Sciences, Tokyo, Japan, [2]Tokyo Institute of Technology, School of Life Science and Technology, Tokyo, Japan

Transposable elements (TEs) are silenced by epigenetic mechanisms such as DNA cytosine methylation. In plants, DNA methylation in non-CG contexts (mCH), which also represses TEs, is localized only in TEs, not in genes. However, the mechanism of TE-specific mCH establishment is obscure. We previously examined the dynamics of TE-specific mCH establishment by restoring mCH machinery in the generation following global loss of mCH. The results showed that mCH can be established by the mechanism independent of RNAi but dependent on mCG. They also suggested the mechanism involves the histone H2A variants H2A.W and H2A.Z. H2A.W, close to macroH2A in animals, is known to accumulate in TEs while H2A.Z accumulates in genes. To investigate their contribution to TE-specific mCH establishment, we genetically examined the dynamics of mCH establishment in each H2A variant mutation background. Our results revealed that H2A.W positively affects mCH establishment while H2A.Z negatively affects it. Moreover, ChIP-seq experiments show that simultaneous loss of mCH with H2A.Z or H2A.W caused the ectopic localization of H2A.W or H2A.Z, respectively, and these mis-localization is highly correlated with the mCH establishment dynamics. In addition, simultaneous loss of mCH with mCG disturbed the distribution patterns of H2A.W and H2A.Z, but their distribution are still anticorrelated. The presence of H2A.W and the absence of H2A.Z is highly correlated with the level of mCH establishment. Thus, our results suggest that H2A.W promotes mCH establishment while H2A.Z inhibits it. Importantly, the mutation alone in the H2A variants did not alter mCH distribution, indicating that they are dispensable for mCH maintenance but are required for mCH establishment. In addition, mCH restoring in the absence of H2A.Z caused the ectopic mCH establishment in a subset of genes. This ectopic mCH establishment could be due to the ambiguous border between TEs and genes caused by the loss of H2A.Z, as it should normally accumulates in genes. Taken together, our results indicated that the appropriate localization of H2A variants contributes to the differentiation between TEs and genes and the formation of TE-specific mCH patterns.

HUMAN DNA LIGASE 1 FACILITATES LINE-1 RETROTRANSPOSITION

Ryuto Ohara[1,2], Shintaro Adachi[1,2,3], Yuzo Watanabe[4], Fuyuki Ishikawa[2,3], Tomoichiro Miyoshi[1,2,3]

[1]RIKEN, IMS, Yokohama, Japan, [2]Kyoto Univ., Gene Mechanisms, Kyoto, Japan, [3]Kyoto Univ., Radiation Biology Center, Kyoto, Japan, [4]Kyoto Univ., Proteomics Facility, Kyoto, Japan

Transposable elements comprise approximately 45% of the human genome. Long INterspersed Element-1 (LINE-1 or L1) is the only active and autonomous retrotransposon in the genome and it mobilizes via a copy-and-paste mechanism termed retrotransposition. Althogh L1 retrotransposition contributes to genetic diversity within the human population, it also causes several diseases, such as cancer. L1 encodes two proteins: ORF1p and ORF2p. ORF1p is an RNA-binding protein, whereas ORF2p possesses endonuclease (EN) and reverse transcriptase (RT) activities. After ORF1p and ORF2p are translated, these L1 proteins bind to L1 RNA, which encodes itself, and form the ribonucleoprotein particles (RNPs) in the cytoplasm. After the L1 RNP is imported into the nucleus, ORF2p digests genomic DNA and reverse transcribes the L1 cDNA with its EN and RT activities. While the activities of these L1 proteins are required for L1 retrotransposition, the exact mechanism by which L1 cDNA, reverse transcribed by ORF2p, is inserted into the genome remains unclear.

In eukaryotes, DNA end joining is facilitated by three types of DNA ligases: DNA ligase 1, 3, and 4 (LIG1, LIG3, and LIG4). To examine which DNA ligases are required for L1 retrotransposition, we measured the L1 retrotransposition frequency under each DNA ligase knockdown condition. Our results revealed that only LIG1 knockdown, but not the other DNA ligases, resulted in a reduction of L1 retrotransposition frequency to approximately 40% compared to the control knockdown. Furthermore, we also generated a LIG1 knock-out cell line, which showed the frequency of approximately 20% to the control cells. These data suggest that LIG1 facilitates L1 retrotransposition by joining L1 cDNA to the genomic DNA.

IDENTIFYING NOVEL REGULATORS OF LINE-1 EXPRESSION THROUGH CRISPR/CAS9 SCREENS

Ozgur Oksuz, Simon Chu, Cedric Arisdakessian, Liyang Diao, Dennis Zaller, Kim Long, Heike Keilhack, Sarah Knutson

ROME Therapeutics, Biology, Boston, MA

Long Interspersed Elements-1 (LINE-1) are transposable elements that make up roughly 17% of the human genome. These elements can copy and insert themselves into new genomic locations. Typically, they are kept silent in healthy tissues but are expressed in various human diseases. LINE-1 expression has been associated with aging, neurodegenerative disorders, cancer, and autoimmune diseases. Despite the correlation of LINE-1 expression with disease, little is understood of how LINE-1 expression is regulated. To explore this, we developed a cellular reporter system to monitor the protein levels of LINE-1 encoded ORF1p and ORF2p simultaneously. Using genome-wide CRISPR/Cas9-based screens with this reporter system, we identified genes that control LINE-1 expression at both the RNA and protein levels. Besides known regulators like the HUSH complex, our screening uncovered previously unknown regulators of ORF1p and ORF2p, many of which appear to be involved in key molecular pathways implicated in human disease. These findings may enhance our understanding of the molecular mechanisms regulating LINE-1 and provide insights into potential therapeutic targets for diseases linked to LINE-1 dysregulation.

THE TRANSCRIPTIONAL REPERTOIRE OF TRANSPOSABLE ELEMENTS DURING MAMMALIAN PREIMPLANTATION DEVELOPMENT

Marlies E Oomen[1], Diego Rodriguez-Terrones[1], Mayuko Kurome[2], Valeri Zakhartchenko[2], Lorenza Mottes[1], Kilian Simmet[2], Camille Noll[1], Tsunetoshi Nakatani[1], Carlos Michel Mourra-Diaz[1], Irene Aksoy[3], Pierre Savatier[3], Jonathan Göke[4], Eckhard Wolf[2], Henrik Kaessmann[5], Maria-Elena Torres-Padilla[1,6]

[1]Helmholtz Zentrum, Institute of Epigenetics and Stem Cells, München, Germany, [2]Ludwig-Maximilians-Universität, Genzentrum, München, Germany, [3]Université Lyon 1, INSERM, Lyon, France, [4]Agency for Science, Technology and Research, Genome Institute of Singapore, Singapore, Singapore, [5]Heidelberg University, Center for Molecular Biology, Heidelberg, Germany, [6]Ludwig-Maximilians Universität, Faculty of Biology, München, Germany

Transcriptional activation of the embryonic genome (EGA) is a major event that enables the embryo to become independent from maternal control. Along with single-copy genes, transposable elements (TEs) also become activated during this developmental time. Although the importance of TE expression has been shown for several individual elements in mouse and human, the full extent of TE activation and its role during early mammalian development remain unknown. Here, we have performed a comprehensive genome-wide analysis of TE expression across five mammalian species throughout preimplantation development. Using a novel protocol that enables capture of 5' transcript information, we provide evidence for autonomous TE-driven expression, thus indicating independent transcriptional regulation from the host genome. We document extensive transcriptional rewiring of LINEs, SINEs and LTRs at and around mammalian EGA. We identify TEs with similar expression dynamics across species, suggesting shared regulation of a specific TE repertoire, and highlight species-specific patterns. Our work highlights evolutionary conservation of TEs, as well as their transcriptional activity, and provides a powerful resource for understanding their co-option during mammalian development.

TRANSPOSON-MEDIATED EVOLUTION OF BAT IMMUNE SIGNALING RECEPTORS

Andrea D Ordonez, Holly Allen, Edward B Chuong

University of Colorado, Boulder, Molecular, Cellular, Developmental Biology, Boulder, CA

Bats are reservoirs for zoonotic viruses that pose significant threats to human health. There is growing molecular evidence that bats evolved unique immune adaptations allowing them to coexist with viruses. Amongst these, are adaptations predicted to contribute to a subdued inflammatory response in bats, such as a decrease in activity or complete loss of important antiviral sensors and inflammasome proteins.
Alternative splicing of immune genes contributes to novel or regulatory protein function but is a poorly explored area of species-specific immune adaptations. By leveraging long-read RNA sequencing from primary tissues of the Jamaican fruit bat, Artibeus jamaicensis, we identified 63 candidates for bat-specific isoforms of immune genes, where the coding sequence of the alternative transcript is truncated compared to that of the canonical. Of these, 27 isoforms contain a transposon exonization event, where the transposon acts as an alternative promoter, splice site, or poly-A signal mediating the truncation. One of the candidates we are exploring further is a truncated isoform of the interleukin 18 receptor (IL18R1) gene that lacks most of the intracellular signaling domain, referred as IL18R1-short in this investigation. This truncation is mediated by a LINE2 transposable element (L2-TE) that introduces an early poly-A signal. Interestingly, IL18R1-short (and the L2) is conserved in humans but very lowly expressed in only a few tissues. Comparatively, bat IL18R1-short is highly detected in all Artibeus tissues tested thus far. This suggests a bat-specific adaptation that increases the relative expression of this truncated isoform. To investigate the mechanism resulting in the differential expression of IL18R1-short in bats and humans, we are studying the relative strengths of the poly-A signal from the L2 element, and splicing regulation for the terminal exons of the human and bat IL18R1 genes. Moreover, this gene is an interesting candidate because its ligand, IL18, is a pro-inflammatory signal. We predict that IL18R1-short is a decoy receptor and down-regulates IL18 induced inflammatory responses. We are over-expressing IL18R1-short in various human and bat cells to test how it modulates IL18 signaling.
This investigation highlights the role that transposons can have on isoform diversification. Given that different species have their own unique history of transposition events, it implies that transposons can be a strong source of species-specific adaptations. With the advent of long-read sequencing and the lowering cost of this technology, we can explore the role of transposons in isoform diversification across multiple species, to fully understand these mechanisms and assess their impact on transcriptome evolution.

REGULATORY INTERACTIONS BETWEEN 3' tRNA FRAGMENTS AND DOMESTICATED LONG TERMINAL REPEAT RETROTRANSPOSONS

Matthew Peacey, Andrea Schorn

Cold Spring Harbor Laboratory, Gene Regulation & Inheritance, Cold Spring Harbor, NY

3' tRNA fragments (tRF3s) are a class of small RNA derived from the 3' end of transfer RNAs. Through complementarity to the tRNA primer binding sites (PBSs) of long terminal repeat (LTR) retrotransposons, tRF3s restrict the mobility of elements active in mice. This regulation occurs through post-transcriptional silencing or interference with reverse transcription. Although active transposons are detrimental to host fitness, some inactive elements have been co-opted to serve host functions. We have established a computational pipeline to predict novel targets of tRF3s in the mouse and human transcriptomes, including examples of domesticated LTR retrotransposons that carry a conserved PBS. Amongst these is *Peg3*, an imprinted gene encoding a transcription factor derived from Gypsy/Ty3 Gag. Mutation of the tRF3 target site de-represses *Peg3* reporter RNA, consistent with post-transcriptional regulation at this site. Ongoing work aims to demonstrate silencing of *Peg3* by tRF3s in an endogenous context, and to identify further examples of the re-purposing of tRF3s to regulate domesticated LTR-retrotransposons.

HUMAN ENDOGENOUS RETROVIRUSES (HERVs) ASSOCIATED WITH GLIOBLASTOMA RISK AND PROGNOSIS

Harun Mazumder[1], Hui-Yi Lin[1,2], Melody Baddoo[3,4], Wojciech Gałan[5], Diana Polania-Villanueva[6], Chindo Hicks[7], David Otohinoyi[7], Francesca Peruzzi[2,6,8], Zbigniew Madeja[9], Victoria P Belancio[4,10], Erik Flemington[3,4], Krzysztof Reiss[2,6,11], Monika Rak[2,6,9,11]

[1]Louisiana State University Health Sciences Center, School of Public Health, Biostatistics Program, New Orleans, LA, [2]Louisiana Cancer Research Center, ., New Orleans, LA, [3]Tulane University School of Medicine, Department of Pathology & Laboratory Medicine, New Orleans, LA, [4]Tulane Cancer Center, Tulane Health Sciences Center, New Orleans, LA, [5]Independent researcher, ., ., Poland, [6]Louisiana State University Health Sciences Center, Stanley S. Scott Cancer Center, New Orleans, LA, [7]Louisiana State University Health Sciences Center, School of Medicine, Department of Genetics and the Bioinformatics and Genomics Program, New Orleans, LA, [8]Louisiana State University Health Sciences Center, School of Medicine, Department of Medicine, New Orleans, LA, [9]Jagiellonian University in Kraków, Faculty of Biochemistry, Biophysics, and Biotechnology, Department of Cell Biology, Kraków, Poland, [10]Tulane University School of Medicine, Department of Structural and Cellular Biology, New Orleans, LA, [11]Louisiana State University Health Sciences Center, School of Medicine, Department of Interdisciplinary Oncology, New Orleans, LA

Emerging evidence suggests expression from human endogenous retrovirus (HERV) loci likely contributes to, or is a biomarker of, glioblastoma multiforme (GBM) disease progression. However, the relationship between HERV expression and GBM malignant phenotype is unclear. Applying several *in silico* analyses based on data from The Cancer Genome Atlas (TCGA), we derived a locus-specific HERV transcriptome for glioma that revealed 211 HERVs significantly dysregulated in the comparisons of GBM vs. normal brain (NB), GBM vs. low-grade glioma (LGG), and LGG vs. NB. Our analysis supported development of a unique HERV scoring algorithm that segregated GBM, LGG, and NB. Interestingly, lower HERV scores showed correlation with lower survival in GBM. However, HERV scores were less robust in predicting LGG survival or LGG progression to GBM. Functional prediction analysis linked the 211 HERV loci with 18 voltage-gated potassium channel genes. The functional link between dysregulated HERVs and specific potassium channel genes may contribute to better understanding of GBM pathogenesis, disease progression, and possibly drug resistance.

This project is financed by NIH COBRE grant no. 1P20GM121288 (LSUHSC New Orleans Center for Translational Viral Oncology - CTVO).

MASTER REGULATOR TRANSCRIPTION FACTORS AND THEIR REGULATION OF TRANSPOSABLE ELEMENTS IN BURKITT LYMPHOMA AND DIFFUSE LARGE B CELL LYMPHOMA

Helena Reyes-Gopar[1,2], Bhavya Singh[3], Jez L Marston[4], Matthew Greenig[5], Hugo Tovar[2], Enrique Hernández-Lemus[2], Ethel Cesarman[6], Douglas F Nixon[1,4], Matthew L Bendall[4]

[1]Feinstein Institutes for Medical Research, Northwell Health, Institute of Translational Research, Manhasset, NY, [2]Instituto Nacional de Medicina Genómica, Departamento de Genómica Computacional, Mexico City, Mexico, [3]Icahn School of Medicine at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, [4]Weill Cornell Medicine, Division of Infectious Diseases, New York, NY, [5]University of Cambridge, Department of Chemistry, Cambridge, United Kingdom, [6]Weill Cornell Medicine, Department of Pathology and Laboratory Medicine, New York, NY

Transposable elements (TEs) like Human Endogenous Retroviruses (HERVs) and Long Interspersed Nuclear Elements (L1) are involved in human genomic regulation. Although transcribed in lymphomas1, their specific participation in the regulatory network remains poorly understood. We investigated TEs in Burkitt Lymphoma (BL) and Diffuse Large B-Cell Lymphoma (DLBCL) by integrating TE expression at locus resolution to infer their Master Regulators (MRs)— key transcription factors driving the phenotype-specific gene regulatory networks. We analyzed RNA-seq data from 113 BL and 529 DLBCL samples with Telescope2 and constructed lymphoma regulons for the human transcription factors. Then, we identified MRs using healthy germinal center B-cell RNA-seq data as a control. We found distinct MRs driving the transition from healthy to diseased states in BL and DLBCL, with HERV and L1 transcripts within MRs' regulons, highlighting TEs' contribution to oncogenic processes. Despite their shared origin, BL and DLBCL have different top MRs, each with unique TEs in their regulon. TFs known to be important in BL and DLBCL like BLC6, NFkB, MYC, MEF2C, and CREB3L1 were confirmed, with novel HERV and L1 targets. These findings offer new insights into BL and DLBCL molecular mechanisms and potential regulatory relationships involving TEs. Further investigation into these interactions' implications may aid in discovering novel therapeutic targets. Funding: NIH grant CA260691, References: 1. PMID: 37333202, 2. PMID: 31568525

UNDERSTANDING HOW A CONJUGATIVE TRANSPOSON SPREADS ANTIBIOTIC RESISTANCE GENES AMONG GRAM-POSITIVE BACTERIA

Carlos Rojas-Cordova, Máté Fülöp, Julie Righi, Orsolya Barabas
University of Geneva, Department of Molecular and Cellular Biology, Sciences III, Geneva, Switzerland

Conjugative transposons (CTn) are main drivers of antibiotic resistance spread among bacteria. Among them Tn*1549*, found in various Gram-positive bacteria like *E. faecalis*, encodes vancomycin resistance, a last-resort antibiotic. Thus, its spread among different pathogens, especially in hospital settings is of high concern.
Previous *in vivo* studies have shown that Tn*1549* encoded proteins, Int and Xis, are necessary for mobilization, allowing excision from the host genomic DNA. Int is also required for integration in the recipient cell, but the role of Xis here is not completely understood. Moreover, unlike prototypical tyrosine recombinases, Tn*1549* prefers to integrate at AT-rich sites that are not particularly conserved. This questions to what extent Tn*1549* transposition resembles the mechanism of site-specific tyrosine recombination. Moreover, *in vitro* studies had failed so far to reconstitute excision and integration, hampering the structural characterization of transposition intermediates.
Here, we show that Tn*1549* co-opts the host-encoded protein HU (from *E. faecalis*) for excision and integration. This knowledge allowed us to reconstitute an excision (EXC) and integration (INT) intermediate and solve their structures using single-particle cryo-EM at near-atomic resolution. The EXC-model is composed of a four-way DNA junction (HJ-DNA) together with the proteins Int, Xis and HU. It shows a tetrameric assembly of four Int monomers sitting on a HJ-DNA molecule. Each Int subunit interacts with its neighbour through a cyclical exchange of the C-terminal alpha helix, revealing some similarities to the well-known tyrosine recombinases. Moreover, HU and Xis orchestrate specific DNA twists at the transposon ends, allowing Int to bind its core and regulatory DNA sites. On the other side, the INT-model was trapped in a pre-cleavage state, where four Ints are binding two separate DNA molecules. The overall assembly resembles the EXC complex model, but in the INT complex, two non-neighboring integrases show a more compact C-terminal arrangement with all their catalytic residues positioned close enough to the DNA cleavage site, ready to perform the subsequent DNA cleavage and joining reactions, while the other two Int molecules are in an inactive state and have a more extended C-terminal tail.
Altogether, these results exemplify how conjugative transposons have co-evolved with and adapted in their hosts to ensure their proper propagation.

BREAK FREE AND THRIVE: A CASE OF AN ACTIVE ENDOGENOUS RETROVIRAL ELEMENT IN THE SOMA OF *D. MELANOGASTER*

Natalia Rubanova[1,2,3,4], Darshika Singh[5], Louis Barolle[5], Fabienne Chalvet[5], Sophie Netter[5], Mickael Poidevin[5], Nicolas Servant[2], Allison J Bardin[1], Katarzyna (Kasia) Siudeja[5]
[1]Institut Curie, INSERM, Sorbonne Université, CNRS, Génétique et Biologie du Développement, Paris, France, [2]Mines Paris - PSL, Institut Curie, INSERM, Cancer et génome: Bioinformatique, biostatistiques et épidémiologie d'un système complexe, Paris, France, [3]Current affiliation: Boston Children's Hospital, Division of Genetics and Genomics, Boston, MA, [4]Current affiliation: Harvard Medical School, Department of Pediatrics, Boston, MA, [5]I2BC, CEA, Université Paris-Saclay, CNRS, INSERM, U1280, Gif-sur-Yvette, France

Reports from human and model organisms show that some TE loci can be active in healthy and diseased tissues. Their activity can lead to a variety of potentially deleterious consequences through, for example, either genome instability on successful and unsuccessful transposition, or by promoting ectopic recombination; or by triggering immune responses; or by rewiring cellular transcriptional programs. In order to prevent these, host cells have developed various mechanisms to tightly regulate the whole life-cycle of TEs from transcription to integration. Yet, some TE loci are able to evade these mechanisms of repression, but how exactly this happens is not well understood.
It was shown before that somatic retrotransposition in the *D. melanogaster* intestinal tissue can lead to tumor suppressor inactivation and formation of gut neoplasia (Siudeja, van den Beek, 2021). I will present our follow-up study in non-diseased tissue in this model system where we observe active *in vivo* somatic retrotransposition of an endogenous retroviral element *rover*. There are only 19 nearly full-length copies of *rover* elements in the genome of the strain we work with, each of which has accumulated distinct sequence variants. We identified the first somatically active donor locus in *D. melanogaster* by comparing these variants with variants present in the *de novo* insertions and by analyzing per copy expression patterns.

The results of our computational and experimental work show:
(1) that substantial number of retrotransposition events can be detected already in young flies and that there is a moderate increase with age of detected transposition events;
(2) that the activity of the donor locus residing in an intron of an actively transcribed gene in a permissive chromatin environment can be explained neither by sequence variants that could have altered core promoter elements or binding sites of transcription factors active in the intestinal lineage nor by permissive chromatin environment alone;
(3) that the upstream genomic locus modulates the transcriptional activity of the donor locus.

TRANSCRIPTION INITIATION AND ITS REGULATION IN MOUSE L1s

Karabi Saha, Wenfeng An

South Dakota State University, Pharmaceutical Sciences, Brookings, SD

Long interspersed element type 1 (L1) transcription is controlled by an internal promoter contained in the 5' untranslated region (5'UTR) of a full-length element. It has been shown that transcription start sites (TSSs) of human L1 are clustered at the beginning of its 5'UTR and the precision of transcription initiation is controlled by transcription factor YY1. The laboratory mouse is an essential model for investigating L1 retrotransposition and the associated consequences in vivo. However, mouse L1 5'UTRs are distinctly different from human L1's: the former is organized into tandem repeats called monomers and separated from the open reading frames by a nonmonomeric sequence called tether. Unlike the human L1 promoter, transcription initiation and its regulation in mouse L1 subfamilies remain unknown. In this study, we aim to investigate the position and frequency of TSS usage in mouse L1 promoters and the role of YY1 in transcription initiation. A streamlined cloning method was established for cloning different types of mouse L1 promoters. TSSs were determined using RNA ligation-mediated 5' rapid amplification of cDNA ends (5'RACE) incorporating a highly efficient reverse transcriptase enzyme followed by Nanopore sequencing. Our results showed that mouse L1 TSSs are distributive and transcription could be initiated from either monomer or tether sequence. In a two-monomer promoter construct of the mouse L1 Tf_I subfamily, 15% of transcripts were initiated from monomer 2 (M2), 65% from monomer 1 (M1), and 20% from the tether. In M2, 45% of transcripts started from the YY1-binding motif and 51% from a location 28 nt upstream. In contrast, in M1, 32 % of transcripts started from the YY1-binding motif, 7% from 22 nt upstream, and 59 % from 52 nt downstream. In addition, we showed that YY1 transcriptionally activates mouse L1 5'UTR and is responsible for the synergy between the monomers. As transcription initiation is a major determinant for the proportion of active L1 loci in the next round of retrotransposition, these findings have important implications on L1 biology in the mouse model.

CHARACTERIZATION OF THE TN7-LIKE TN6230 ELEMENT THAT DISPLAYS A NOVEL DUAL TNIQ CONFIGURATION

Erica Sawyer, Shravanika Kumaran, Alberto Correa, Richard Schargel, Joseph E Peters

Cornell University, Microbiology, Ithaca, NY

Tn6230 family elements are a distinct group of transposons found on mobile plasmids and at a chromosomal attachment site at the C-terminal coding end of the putative oxidoreductase gene, *yhiN*. These elements are present in diverse microorganisms, including *Salmonella*, *Escherichia coli*, and *Shigella*, and they often contain cargo genes for metal resistance, antibiotic resistance, and other functions. One common cassette of metal resistance may have been exchanged in an agricultural setting from an unrelated group of Integrating and Conjugating elements (ICE). Similar to Tn7-like elements, Tn6230 elements possess a TnsC AAA+ ATPase that likely regulates the interaction between target-site selecting proteins and a heteromeric transposase consisting of a TnsA-family endonuclease and TnsB-family transposase. However, Tn6230-like elements have two TniQ-family proteins hypothesized to mediate target-site selection through an undetermined mechanism. In this work, we characterized a *Serratia marcescens* Tn6230-like element in a heterologous *E. coli* host. First, we demonstrated that the larger TniQ-family protein selectively directs transposition to the target site *yhiN* in a single orientation; conceptually similar to TnsD in prototypic Tn7. Second, we showed that TnsC mediates transposition in a manner distinct from Tn7. Mutations in the ATP-hydrolyzing Walker-B motif result in markedly lower transposition frequency compared to wild-type TnsC. This directly contrasts with the high levels of random transposition seen with prototypic Tn7 Walker-B mutations. Together, these results reveal the unique attributes of Tn6230-like elements and suggest a fundamental mechanistic difference in comparison to Tn7. Preliminary bioinformatic work suggests that Tn6230 coopted the Tns functions independently. Ongoing work will further characterize target site immunity and the role of the second TniQ protein, providing deeper insights into the underlying mechanisms of Tn6230 elements, revealing their functionality in the context of other Tn7-like elements, and expanding our understanding of these diverse elements.

ELUCIDATION OF TRANSPOSON TARGETING MECHANISMS IN TN5053 AND TN7-LIKE ELEMENTS

Richard D Schargel, Joseph Peters

Cornell University, Microbiology, Ithaca, NY

Transposons are mobile genetic elements that can move between positions in a genome. Transposable elements generally mobilize to random or slightly biased positions in a genome. This can result in deleterious mutations when the transposon inserts into an essential gene. The Tn7 and Tn7-like family of transposons avoid this issue by only integrating into specific safe sites within the chromosome. Additionally, most of Tn7-like elements have an alternate targeting pathway that favors insertions into features found on other genetic elements that would facilitate transition to new bacterial hosts. Target site selection in this family is executed by, or in association with, a TniQ family protein in a typically DNA sequence-specific manner. An unusual type of transposon within this family, called Tn5053, carries a TniQ that seemingly targets plasmid dimer resolution sites. Interestingly, Tn5053 also contains a plasmid dimer resolution site alongside its cognate resolvase; TniR. This implies that the Tn5053 encoded resolution site and resolvase compensate for the disrupted target resolution site, rendered dysfunctional during transposition, thereby acting as a form of addiction system. Transposition into plasmid resolution sites is expected to allow for the proliferation of Tn5053 across bacterial hosts, thereby spreading antibiotic and biocide resistance. Tn5053 uses three proteins to catalyze replicative transposition into resolution sites. TnsB is the transposase, responsible for donor element breaking and joining. TnsC, is a AAA+ ATPase that likely functions as a regulator, coordinating the recruitment and activation of the transposase exclusively upon identification of a target site. TniQ is predicted to be the target site selector. Here we investigate the interactions between Tn5053's proteins (TnsB, TnsC, TniQ) and the resolvase protein, and compare the regulatory mechanisms of TnsC with those in Tn7 and Tn7-like elements, including CRISPR-associated transposons (CASTs). Additionally, this project explores the potential of engineering the target-site specificity of Tn5053 and other Tn7-like transposons.

A LINK BETWEEN CONSERVED KRAB ZINC FINGER PROTEINS AND POLYCOMB-MEDIATED GENE REGULATION

Anna D Senft[1], Peter Lindner[1], Nick Miller[1], Shelley N Dolitzky[1], Amanda N Ivanoff[2], Alexander Grinberg[1], Sherry A Ralls[1], Todd S Macfarlan[1]

[1]National Institutes of Health, The Eunice Kennedy Shriver National Institute for Child Health and Development, Bethesda, MD, [2]University of Buffalo, Department of Biological Sciences, Buffalo, NY

Endogenous retroviral sequences (ERVs) influence genomes directly and indirectly. One indirect consequence of vertebrate ERV infiltration is the abundant evolution of KRAB zinc finger proteins (KZFPs) that sequence-specifically regulate ERVs. Like ERVs, KZFPs can be co-opted for host function, yet the vast majority of KZFPs remain unexplored. We describe a deletion of a conserved cluster of seven KZFPs in mice which results in perinatal lethality. The lethality is recapitulated by deleting a single, 320-million-year-old member of the cluster, Zfp777. We identify a ventricular septal defect in hearts of Zfp777 KO embryos which could not only explain the lethality but also why the human ortholog ZNF777 is intolerant to mutations in healthy humans and is - when mutated - associated with a ventricular septal defect. Profiling of Zfp777 genome-wide occupancy in mouse embryonic stem cells identifies lack of localization to ERVs but binding to critical Polycomb-repressed developmental transcription factors, which Zfp777 represses during subsequent development. Lack of typical KZFP-associated repressive mechanisms and loss of Polycomb factors at a subset of target genes in Zfp777 KO mouse embryonic fibroblasts indicate that Zfp777 uses Polycomb-mediated repression. Intriguingly both KZFP- and Polycomb systems are capable of targeting ERVs. We propose that these two ancient repressive systems co-operate to mediate gene regulation in mammals.

THE ROLE OF AGO AND PIWI PROTEINS IN ENDOGENOUS RETROVIRUS CONTROL

Helene Sertznig, Jenna Wilken, Maggie Gravano, Stefan Constantinou, Juanjuan Xie, Andrea J Schorn

Cold Spring Harbor Laboratory, Gene Regulation and Inheritance, Cold Spring Harbor, NY

Mobility of active transposable elements due to epigenetic reprogramming in development and disease is highly mutagenic and threatens genome stability. 3'-tRNA-derived fragments (3'-tRFs) are cleavage products of mature tRNAs and target long terminal repeat (LTR)-retroelements at their highly conserved tRNA primer binding site (PBS). 3'-tRFs are able to silence mobile genetic elements during epigenetic reprogramming in the absence of other small RNAs. Argonaute (AGO/PIWI) proteins are at the core of the RNA interference (RNAi) pathway and bind small RNAs, including 3'-tRFs, to induce post-transcriptional gene silencing.
Here, we identified and characterized RNAi pathway proteins that regulate endogenous retrovirus (ERV) retrotransposition and potentially mediate 3'-tRF silencing. Using highly reproducible and robust retrotransposition assays, our laboratory demonstrated that the highly active ERV family in mouse, ETn/MusD (Early Transposon/Mus Musculus Particle D), is a major target of 3'-tRFs. CRISPR-Cas9-mediated knock-out of AGO/PIWI proteins combined with retrotransposition assays showed that they strongly affect retrotransposition in the ETn/MusD mouse model in human HeLa cells, indicating an involvement in 3'-tRF silencing. Single knockout of AGO1-4 and PIWIL1-4 had distinct effects on retrotransposition, prompting us to investigate their impact on retrotransposition intermediates and 3'-tRF levels in more detail.
We propose that tRNAs are an ancient substrate of the RNAi pathway and 3'-tRFs play an important role in regulating translation and replication of retrotransposons.

TRANSPOSON MOBILITY UNDER HOST-MIMICKING CONDITIONS IN THE PATHOGENIC YEAST *CRYPTOCOCCUS DENEOFORMANS*

Samantha R Shaltz, Asiya Gusa

Duke University, Molecular Genetics and Microbiology, Durham, NC

Cryptococcus species are environmental fungi that grow in the form of encapsulated yeast and can produce spores capable of infecting humans. To survive the environment-to-host transition, *Cryptococcus* must overcome various stressors (e.g., limited nutrients, elevated temperatures, increased CO_2, and changes in pH). In previous studies with *Cryptococcus deneoformans*, growth under heat stress at human body temperature (37ºC) led to mutations by transposable elements that caused drug resistance to 5-fluoroorotic acid (5FOA) and other antifungals (Gusa and Williams et al, 2020). These studies were performed in nutrient-rich media (YPD) and did not mimic the mammalian host environment. Under these conditions, TE mobilization into target genes was observed with an increased rate of mutagenesis at higher temperatures. Movement of two transposons was detected; the T1 DNA transposon and the Tcn12 LTR-retrotransposon. Using a murine model of infection, these same TEs were captured moving into target genes. In this study, to better simulate host conditions, *C. deneoformans* was incubated at 30º and 37º in nutrient-limiting tissue culture media (RPMI), with and without 5% CO_2. Colonies growing on drug plates (5FOA) were screened for mutations at specific target genes, *URA3* and *URA5*. We saw elevated movement of the LTR-retrotransposon, Tcn12, into our reporter genes when grown under heat stress, though the addition of CO_2 did not significantly increase the proportion of insertions. Experiments are underway to assess differences in mutation rates. Additionally, to better understand the mechanisms of stress-induced mobility, we are constructing transposon reporter systems in different genetic backgrounds. We hope to identify different environmental stressors that stimulate TE mobility and potentially contribute to genetic adaptation during infection.

MAMMALIAN-SPECIFIC MIR RETROTRANSPOSONS RECRUIT ACTIVATING HISTONE MARKS ON ENHANCERS AND PROMOTERS OF CODING GENES AS STEM CELLS DIFFERENTIATE

Grace E Smith[1], Aristeidis G Telonis[2,3]

[1]University of Miami Miller School of Medicine, Human Genetics and Genomics program, Miami, FL, [2]University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Miami, FL, [3]University of Miami Miller School of Medicine, Department of Biochemistry & Molecular Biology, Miami, FL

Human stem cells express genes that are dense in primate-specific Alu and mammalian interspersed repeats (MIRs) but as they differentiate, they start expressing genes sparse in Alu but dense in MIR elements. In addition, MIR elements in the promoters or within genes have been implicated in the epigenetic regulation of the corresponding genes. However, a systemic evaluation of the deposition of epigenetic marks on MIR and Alu retrotransposons during differentiation is lacking. Here, we mined chromatin immunoprecipitation followed by sequencing (ChIP-seq) data of four histone marks, H3K27ac, H3K27me3, H3K4me1 and K3K4me3, in human undifferentiated embryonic stem cells (ESC) as well as in ESC-derived differentiated lineages. As expected, there was a greater number of histone peaks in differentiated cells with limited overlap across different lineages or cell types. We examined the location of these marks with respect to Alu and MIR elements on the human genome and found that these histone marks, particularly H3K4me1, were significantly enriched in MIRs in differentiated cells but depleted from Alu elements. Looking at the promoters of the genes that were expressed, those that were marked by H3K27ac or H3K4me3 were depleted from Alu elements but were significantly enriched in MIR elements. Finally, we distinguished the genomic locations where the same histone mark is consistently found across all differentiated cells and those locations where cell-type-specific histone marks appear. We found that the regions with activating marks across all cell types were significantly enriched in both Alu and MIR elements. The genomic regions with cell-type-specific histone marks had no significant biases in MIR density but there was enrichment of Alu elements in certain cell types in H3K27me3 histone marks. Overall, our results suggest that the distribution of activating histone marks on the human genome as stem cells differentiate may be driven by MIR and Alu elements: MIRs attracting chromatin modifying enzymes while Alu elements repelling them. Our findings shed light on a complex interplay between repeat elements on the human genome with the epigenome during normal differentiation and argue for the importance of genome architecture in our understanding of developmental processes.

DNA TRANSPOSONS MINING AND ENGINEERING FOR EFFICIENT CAR-T MODIFICATION

Chengyi Song[1], Bo Gao[1], Yali Wang[1], Saisai Wang[1], Pingjing Zhang[2], Shasha Shi[1]

[1]Yangzhou University, College of Animal Science and Technology, Yangzhou, China, [2]Shanghai Cell Therapy Group Co., Ltd, Cell Therapy, Shanghai, China

Cut-and-paste DNA transposons are widely recognized as natural and stable non-viral gene delivery vectors that offer several advantages, including low cost, low immunogenicity, low cell toxicity, large cargo capacity, and reduced susceptibility to immune system silencing. These transposons have found broad applications in gene therapy, transgenesis, and mutagenesis studies. After conducting large-scale genomic data mining and activity screening using a cell-based transposition assay, we identified over 4,000 cut-and-paste DNA transposons from more than 20 families, belonging to seven superfamilies (hAT, piggyBac, PHIS, Tc1/Mariner, DD82E/Sailor, pogo, Gambor/Hiker). Based on sequence analysis and K divergence of the selected transposons, we submitted 5 Spy/PHIS, 3 PiggyBac, 9 hAT, 3 pogo, one Mariner/Tc1, and 3 DD35E/Hiker for transposition activity evaluation. Among them, we identified 10 highly active transposons: PS from pogo, ZB from Tc1/Mariner, miCN and lrCN from Cleaner/hAT, anBuster from Buster/hAT, PB13, PB84, and PB108 from PiggyBac, and cgSpy and cvSpy from Spy/PHIS. We further characterized the cargo capacity, overproduction inhibition (OPI), and insertion preferences of PS, ZB, and cgSpy. Furthermore, we achieved significant improvements in transposition efficiency and cargo capacity by further engineering and optimizing the PS transposon systems. Additionally, the enhanced PS transposon system, in combination with the Tini-plasmid technique (small vector backbone, approximately 500 bp), significantly improves the efficacy of CAR-T cells engineered even at low doses, resulting in impressive therapeutic outcomes in solid tumors. This evidence underscores the promise of the optimized JL transposase, paired with the condensed donor vector, as an efficacious and secure strategy for gene therapy in clinical practice.

LINE-1 PRODUCES CYTOSOLIC DNA PRODUCTS FROM SELF AND NON-SELF RNAs

Jared Steranka, Ozgur Oksuz, Oluchi Ezekwenna, Wilson McKerrow, Simon Chu, Liyang Diao, Menachem Fromer, Jess Floro, Dennis Zaller, Wenyan Miao, Kim Long, Sarah Knutson, Heike Keilhack

Rome Therapeutics, Biology, Boston, MA

Long Interspersed Element-1 (LINE-1) retrotransposons are genetic parasites that self-propagate through a 'copy and paste' mechanism driven by the LINE-1 encoded open reading frame 2 protein (ORF2p) that contains both endonuclease and reverse transcriptase (RT) domains. The RT activity of ORF2p has been linked to the development of cancer, autoimmune diseases, aging, and neurodegenerative disorders, making ORF2p a potential therapeutic target. ROME is currently developing novel LINE-1 reverse transcriptase inhibitors (RTIs) for therapeutic use. To date, efforts to explore LINE-1 as a therapeutic target have been hindered by a limited mechanistic understanding of LINE-1 RT activity in cells, particularly regarding the diversity of potential cytosolic reverse transcription products generated by LINE-1 RT using both its own RNA and other non-LINE-1 RNA templates. To address this, we have developed methods to identify and quantify LINE-1-mediated cytosolic DNAs, including DNA:RNA hybrids, from cells. These methods enabled us to demonstrate that LINE-1 generates cytosolic DNA products of varying lengths from its own RNA and non-LINE RNAs, both of which depend on LINE-1 RT activity. Treatment with ROME RTIs reduced the abundance and size of these products. These findings provide key mechanistic insights into how LINE-1 reverse transcribes itself and non-LINE-1 RNAs, advancing the development of potential LINE-1 RT inhibitors as therapeutic candidates.

TOWARDS A PANTEOME OF *ARABIDOPSIS THALIANA*

Svitlana Sushko, Zhigui Bao, Fernando A Rabanal, Detlef Weigel

Max Planck Institute for Biology Tübingen, Molecular Biology, Tübingen, Germany

Transposable elements (TEs) are not only selfish genetic elements, but they also drive structural changes in genomic sequences and can produce notable effects on phenotypes. Many advancements in the understanding of TE biology have been made using short-read datasets and comparisons with single reference genomes, but they have mostly focused on presence/absence variation and less so on small-scale mutations in individual TE insertions. The drop on long-read sequencing costs now allows us to completely assemble multiple genomes, shedding light on previously understudied regions of the genomes, especially those that are highly repetitive and dense in TEs. We are generating a new pangenome-scale annotation for a large number of *Arabidopsis thaliana* genome assemblies. This new annotation allows us to characterize TE content and its variation in *A. thaliana* genomes in an unbiased manner. We found the TE content in *A. thaliana* to be highly homogeneous among 182 accessions. It explains a relatively small (~4%) proportion of genome size variation, and most TEs in this species are young, which is consistent with a high TE turnover. The TE composition of the genomes is not explained by the phylogenetic grouping of accessions, although some phylogenetic groups differ in overall TE content. Most *A. thaliana* TE families are present in the majority of accessions, but a few are restricted to smaller sets of accessions. By taking into account a large number of accessions in the species, we found new TE families present in non-reference accessions and explored the TE diversity within *A. thaliana*. Ultimately, the high quality of genome assemblies, which allows for a more precise mapping, together with a pangenome-scale TE annotation will allow us to describe the mutational processes taking place inside TE insertions.

THE ROLE OF RETROTRANSPOSABLE ELEMENTS IN NEURODEVELOPMENT

Mary J Talley[1], Michelle Longworth[1,2]

[1]Lerner Research Institute, Cleveland Clinic, Inflammation and Immunity, Cleveland, OH, [2]Case Western Reserve University School of Medicine, Cleveland Clinic Lerner College of Medicine, Cleveland, OH

Microcephaly is a neurodevelopmental disorder where the head size of the patient is 2-3 standard deviations below average. Recently, we investigated a *Drosophila* model of microcephaly caused by condensin insufficiency. We discovered that increased retrotransposable element (RTE) activity was responsible for the observed increase in neural stem and precursor cells (NSPCs) cell death, which is ultimately responsible for the microcephaly phenotype. Excitingly, microcephaly was rescued in condensin insufficient *Drosophila* when RTE activity was prevented using Nucleoside Reverse Transcriptase Inhibitors (NRTIs). While high levels of RTE activity have been associated with several neurodevelopmental disorders, the baseline RTE activity in the brain has been demonstrated to be higher than other tissues. It is currently unclear whether this high baseline RTE activity is required for proper brain development and how RTE activity above this baseline could contribute to neurodevelopmental disorders. Our goal is to determine the function of RTEs during normal brain development and why increased RTE activity can cause microcephaly. We have knocked down Loki, the *Drosophila* Chk2 homolog, in our condensin insufficient model in order to test if DNA damage response pathways contribute to the phenotype. We found that Loki knockdown partially rescued the microcephaly. This suggests that the cellular response to DNA damage, likely caused by the increased RTE activity, is responsible, at least in part, for causing microcephaly in our condensin insufficient model. Activation of the immune system in response to RTE activation may also be partially responsible for the cell death observed in our *Drosophila* model. Further, we have begun investigating additional microcephaly models, including Citron Kinase and ASPM mutants, to determine if increased RTE activity is seen in other microcephaly models. By understanding the role of RTEs in normal brain development and identifying the mechanisms by which increased RTE activity can cause microcephaly, information on the viability of NRTIs as a treatment for microcephaly, and potentially other neurodevelopmental disorders, will be uncovered.

LINE-1 PROMOTES CHROMOSOMAL TRANSLOCATIONS VIA A REVERSE TRANSCRIPTASE-DEPENDENT MECHANISM

Jianli Tao[1], Luca Alessandri[1], Carlos Mendez-Dorantes[2], Kathleen H Burns[2], Qi Wang[1], Roberto Chiarle[1,3]

[1]Boston Children's Hospital and Harvard Medical School, Department of Pathology, Boston, MA, [2]Dana-Farber Cancer Institute and Harvard Medical School, Department of Pathology, Boston, MA, [3]University of Torino, Department of Molecular Biotechnology and Health Sciences, Torino, Italy

In many cancers, chromosomal translocations are pathogenic events initiated by the generation of DNA double-strand breaks (DSBs). Although LINE-1 (L1) retrotransposons have been implicated in the formation of chromosomal rearrangements detected in cancer genomes, the frequency and mechanisms by which L1 is involved in these rearrangements, including chromosomal translocations, remain poorly characterized and possibly underestimated. By combining high-throughput genome-wide translocation sequencing (HTGTS) and two sequencing techniques to detect L1 insertion sites (PolyA-seq and CELTICS-seq), here we demonstrate that L1 expression significantly increases the frequency of chromosomal translocations at genomic hotspots. We find that L1-induced translocation hotspots are dependent on the L1 reverse transcriptase (RT) activity but largely independent of the L1 endonuclease (EN) activity. These hotspots are skewed towards the transcription end sites (TES) of expressed genes, accessible chromatin regions, and early replicating regions. These data suggest a model wherein L1 RT promotes the resolution of distant breaks, thereby facilitating the formation of chromosomal translocations even in the absence of L1 sequence insertions. Therefore, while cancer genome analyses provide a catalog of L1-induced chromosomal rearrangements based on L1 integrations, our studies suggest that L1 expression may affect a broader landscape of chromosomal translocations, facilitating repair of EN-independent DSBs rather than by causing breaks. This expands our knowledge of the mechanism by which L1 retrotransposons can cause genomic instability in human cancers.

CANONICAL AND NON-CANONICAL CUTTING ACTIVITIES OF LINE-1 ENDONUCLEASE

Benedict Smail[1], Bryant D Miller[1], Trevor van Eeuwen[2], Allison O'Brien[2], Kathleen H Burns[1,3], Martin S Taylor[3,4]

[1]Dana Farber Cancer Institute, Pathology, Boston, MA, [2]Rockefeller University, Laboratory of Cellular and Structural Biology, New York, NY, [3]Mass General Brigham and Harvard Medical School, Pathology, Boston, MA, [4]Brown University, Pathology and Laboratory Medicine, Providence, RI

The apurinic/apyrimidinic (APE)-like endonuclease (EN) of the long interspersed element 1 (LINE-1, L1) initiates retrotransposition by nicking double-strand DNA at a consensus 5′-TTTT↓AA sequence, creating a 3′-OH group that primes L1 cDNA polymerization through a mechanism termed “target primed reverse transcription” (TPRT). The consensus motif reflects the site of completed retrotransposition events, which depend on A-T base pairing between the cleaved genomic DNA and the L1 RNA poly(A) tail. Therefore, the consensus may not represent intrinsic properties of L1 EN, and it remains unknown whether L1 EN also cuts other sequences. To investigate both canonical and non-canonical cutting activities of L1 EN, we purified monodisperse WT and catalytic dead E43S-D145N double mutant enzymes. We have verified conventional nicking activity on a circular plasmid, and demonstrated that the E43S, D145N double mutant ablates cutting activity as expected, however denaturing agarose gels reveal substantially less specificity than expected. Investigating this discrepancy, we discovered a robust, non-canonical cutting activity of L1 EN. Secondary structure analysis and cutting activity on a variety of oligonucleotide fragments reveals that L1 EN cuts at positions adjacent to mismatched duplex DNA faster and more specifically than the canonical sequence. In addition, the rate and specificity of both mismatch and canonical cutting increases dramatically near a DNA end including a 5’ overhang, 3’ overhang, bubble, or blunt end with little sequence specificity. Unexpectedly, the 5’ overhang does not have as strong an effect in our assays as was recently reported in radioactive TPRT assays. The effects of a mismatch and proximity to an end act synergistically to increase the rate of reaction. Computational modeling reveals that mismatches may pre-configure the substrate into a conformation mimicking the crystallized transition state. In sum, this data shows that L1 EN cut rate and specificity is determined primarily by substrate secondary structure rather than sequence. This activity may be responsible for cleaving intermediate steps in the complex insertion mechanism and provides support for the hypothesis that L1 could preferentially cut and insert at Okazaki fragments during DNA replication.

MATERNAL piRNAs GUIDE piRNA CLUSTER ADAPTATION TO SILENCE NOVEL TE IN DROSOPHILA

Amna Asif-Laidin, Karine Casier, Zoheir Ziriat, Antoine Boivin, Valérie Delmarre, Stéphane Ronsseray, Clément Carré, Laure Teysset

Sorbonne Université, CNRS, Institut de Biologie Paris Seine, LBD, Paris, France

Silencing of Transposable Elements (TEs) in Drosophila female gonads is achieved by PIWI-interacting RNAs (piRNAs), a class of small RNAs synthesized by heterochromatic loci, called piRNA clusters. These clusters are mostly long and enriched in fragments of TEs that originated from past insertion of mobile TEs. Thus, piRNA clusters can be seen as libraries of mobiles sequences that need to be silenced. The germline piRNA clusters are redefined and activated at each generation thanks to inheritance of maternal piRNAs. Horizontal transfer (HT) of TEs into naïve genomes poses a threat for the host genome integrity due to the absence of inherited complementary piRNAs. The insertion of a newly HT TE into the germline genome is necessary to synthesis specific piRNAs that can trigger silencing, ensuring the maintenance of this new TE in future generations.
Using TE-derived transgenes inserted into subtelomeric germline piRNA clusters and Drosophila genetic techniques, we have investigated the timing of new piRNA appearance to repress novel TE invaders in the generation following TE insertions into a piRNA cluster. By measuring new piRNA synthesis, our results indicate that piRNA clusters can co-opt sequences not present in the maternal pool to synthesize new piRNAs within a few generations. We will present data showing that the distribution profiles of piRNAs across piRNA cluster loci are unexpectedly heterogeneous.

REPURPOSING NATIVE GUIDE RNA-DIRECTED TRANSPOSITION SYSTEMS FOR GENOME ENGINEERING IN EUKARYOTIC ORGANISMS

Jordan Thesier, Robert Wingo, Shan-Chi Hsieh, Michael Petassi, Saadlee Shehreen, Erica Sawyer, Lille Cunic

Cornell University, Department of Microbiology, Ithaca, NY

Transposon Tn7 is renowned for its ability to transpose in a sequence-specific manner into a safe site in the bacterial genome. The Tn7 transposase is formed by two proteins, TnsA and TnsB, and utilizes a AAA+ ATPase, TnsC, to communicate with target-selecting proteins like TniQ or a TniQ-domain containing protein. Only when a target is identified will the element transpose. This control over target site selection has been naturally repurposed by Tn7-like transposons on multiple occasions. In some exciting examples, Tn7-like elements have co-opted CRISPR-Cas systems for RNA-directed transposition, where nuclease-deficient CRISPR-Cas effectors guided by RNAs communicate with transposon machinery to precisely integrate DNA into the recognized targets. CRISPR-associated transposons (CASTs) are attractive as genome-editing tools due to their precise programmable targeting, strict insertion orientation control, large cargo capacity, and independence from homology-directed repair. In this study, we investigate type I-F3 CASTs, transposons that have co-opted the I-F1 CRSIPR-Cas effector, Cascade, for targeting through TniQ.

Type I-F3 CASTs are one of the largest group of CASTs in sequenced genomes. We developed a pipeline to explore the diversity of I-F3 CASTs and identified ~40 orthologs. By optimizing core transposon machinery with nuclear localization tags for each system, we discovered a gain-of-function TnsC mutation that facilitates TnsABC-only transposition without the targeting TniQ/Cascade units. Testing these I-F3 orthologs in *E. coli* with a minimal TnsABC* system, we found that nearly all could be re-established. We further analyzed a highly active subset of these orthologs with targeting components, observing that most chosen CASTs were able to respond to TniQ/Cascade, and identified ten systems more efficient than the most used systems. To address the activity of these CASTs in eukaryotes, we developed a system in *S. cerevisiae* to detect donor element excision and did indeed observe transposition with a minimal TnsABC* system. We plan to test our optimized targeted transposition systems with TniQ/Cascade in yeast, as well as diverse eukaryotic systems.

This work underscores the potential of I-F3 CASTs to enhance genome engineering and establishes a strong foundation for further testing in eukaryotes.

MODELLING PROLIFERATION OF RETROTRANSPOSONS IN SOMATIC STEM CELL LINEAGES OF PLANTS

Sou Tomimoto[1,2], Eriko Sasaki[1], Akiko Satake[1]

[1]Kyushu University, Department of Biology, Fukuoka, Japan, [2]Kyushu University, Graduate School of Systems Life Sciences, Fukuoka, Japan

Recent advancements in sequencing technology have enabled the quantitative identification of somatic mutation accumulation in long-lived organisms. This has significant implications especially in plants, where somatic mutations can be inherited and influence evolutionary trajectories. While extensive research has shed light on single nucleotide variations (SNVs) and their age-dependent linear accumulation, as observed in trees, the behavior of transposable elements (TEs) as potentially major contributors to somatic mutations remains poorly understood.
To address this, we built simple mathematical models to describe the proliferation process of retrotransposons in somatic stem cell lineages, considering two scenarios: proliferation without TE silencing (Model I) and with silencing upon exceeding a threshold (Model II). In Model I, the TE proliferation process begins with the activation of a single TE at a rate μ. The activated TE then makes own copies at a rate λ, leading to an exponential increase in TE copies as a host plant ages. We derived the expected value and variance of the TE copy number, extending the Yule process. In Model II, TE proliferation is suppressed when the copy number exceeds a maximum threshold m_{max}. Based on these models, we calculated the genetic differences resulting from TEs between two cell lineages that bifurcate from a common ancestral lineage, such as two branch tips of a tree or two clonal ramets of a single genet.
Unlike SNVs, where mutations accumulate linearly, the ages of the cell lineages until and after bifurcation significantly influence the genetic differences due to the exponential nature of TE proliferation. Copy number-dependent TE silencing also yields intriguing patterns, indeed, a high proliferation rate can lead to self-suppression before bifurcation, resulting in no genetic differences. The models provide a framework for understanding the forthcoming genomic studies on TE proliferation in somatic cell lineages, highlighting the complex and fascinating nature of TEs.

GENOMIC ANALYSIS OF SILKWORM *W1 PND* STRAIN

Takuya Tsubota, Kakeru Yokoi

National Agriculture and Food Research Organization, Institute of Agrobiological Sciences, 1-2 Owashi, Tsukuba, Ibaraki, Japan

Silkworm is a lepidopteran model insect that has been utilized for the basic researches as well as for the industrial application, including protein production, silk engineering and others. *w1 pnd* is a strain that is frequently utilized for the transgenic and genome editing study, having white-egg, white-eye and non-diapause traits. Recently we carried out the time-course transcriptomic analysis of the silk gland in this strain and obtained a basic data for the entomological as well as medical research. In order to obtain the genomic information of this strain, we here carried out the whole genome sequencing analysis of this strain. The genome was extracted from one female and was applied for the long read genome sequencing using PacBio revio system. This resulted in the production of 86.4 Gb HiFi reads, which corresponds to the x200 coverage of the silkworm genome. These reads were assembled using hifiasm and we successfully established high-quality assemble with N50 of 16Mb. We analyzed each haplotype sequence by using the output data of the assemble and found that there existed a number of structural variants. Most of them corresponded to transposons, suggesting that these transposons are highly active in the *w1 pnd* strain. We are now ongoing the establishment of gene model using transcriptomic data, and we expect that these information will be of use for the efficient transgenic and genome editing study.

CHARACTERZING STRUCTURAL DETERMINANTS OF CYTOPLASMIC LINE-1 ACTIVITY

Trevor van Eeuwen[1], Sidae Lee[2], Bryant Miller[3], Kazuma Kondo[4], Tyler Owens[5], Alan Lambowitz[2], Martin S Taylor[5,6], Michael P Rout[1]

[1]Rockefeller University, Laboratory of Cellular and Structural Biology, New York, NY, [2] The University of Texas at Austin, Department of Molecular Biosciences, Austin, TX, [3]Dana Farber Cancer Institute and Harvard Medical School, Department of Pathology, Boston, MA, [4]Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, Boston, MA, [5]Massachusetts General Hospital and Harvard Medical School, Department of Pathology, Boston, MA, [6]Warren Alpert Medical School, Brown University, Pathology and Laboratory Medicine, Providence, RI

Retrotransposable elements (RTEs), genetic sequences capable of "copy-and-paste" self-replication, comprise approximately 40% of the human genome, with Long INterspersed Element 1 (LINE-1/L1) representing the largest and most active class of RTEs. RTEs are typically suppressed in healthy tissues by several mechanisms, but cells can ultimately lose control over their RTEs because of epigenetic reprogramming associated with cellular senescence or cancer development. This dysregulation of RTEs has been correlated with interferon type-I (IFN-I) production. We have recently demonstrated that L1 expressed in cells can produce cytoplasmic DNA:RNA hybrids which have the potential to behave like pathogen-associated molecular patterns (PAMPs) that stimulate pattern recognition receptors (PRRs) and stimulate innate immune signaling. Consistent with previous studies, we've also observed L1 derepression was correlated with IFN-I signaling and have determined this occurs through cGAS/STING. These data strongly indicate that L1 directly stimulates innate immunity in senescent cells by generating cytoplasmic PAMPS and this is a fundamental process in cellular aging. L1 is capable of priming using self-annealing RNAs (snapback) like those generated by the Alu transcripts, and these may be a source of cytoplasmic DNA:RNA hybrids. We demonstrate here that L1 can prime using RNA and DNA substrates with significantly fewer base pairs than previously thought. Additionally, we demonstrate the structural basis of L1 snapback priming to generate a model for generation of L1 cytoplasmic activity. These findings have potential applications in mediating RT dependent sterile inflammation observed in diseases of aging, which may also be a mechanism of cancer immune evasion.

ACTIVITY ACROSS DIVERSE EUKARYOTIC R2 RETROELEMENT PROTEINS

Briana Van Treeck[1], Nozhat T Hassan[2], David L Adelson[2], Kathleen Collins[1]

[1]University of California, Molecular and Cell Biology, Berkeley, CA, [2]University of Adelaide, School of Biological Sciences, Adelaide, Australia

Non-long terminal repeat (non-LTR) retrotransposons are ubiquitous components of eukaryotic genomes that spread to new sites via target-primed reverse transcription (TPRT) of retroelement RNA by a retroelement encoded protein. In TPRT, reverse transcription is primed from a nick at the DNA target site and cDNA is synthesized directly into the genome. The R2 retrotransposon has served as the model for TPRT mechanism due to its exquisite targeting specificity to a highly conserved sequence within 28S ribosomal RNA genes (rDNA). The sole R2 protein (R2p) has reverse transcriptase and nicking endonuclease activities necessary and sufficient for TPRT. Even though R2 retrotransposons have been effective hitchhikers in a wide diversity of eukaryotic genomes for hundreds of millions of years, biochemical investigations of R2p have almost exclusively used the silkmoth protein, excepting our recent characterization of avian proteins (1). Starting with the expression and purification of R2p from diverse species, I compared their properties including endonuclease and reverse transcriptase activities as well as specificity of RNA template use. R2p from different species exhibit marked diversity in the activities of their various domains, the cross-coordination of these activities, and their specificity in RNA template utilization. This work greatly expands the phylogenetic characterization of active R2p and informs both the unique biology and the potential clinical applications of R2 retrotransposon machinery.

(1) Zhang X, Van Treeck B, Horton C, McIntyre J, Palm S, Shumate J, Collins K. 2024. Harnessing eukaryotic retroelement proteins for transgene insertion into human safe-harbor loci. Nat Biotechnol. https://doi.org/10.1038/s41587-024-02137-y

TN*7* SPREAD AMONG *ESCHERICHIA COLI* STRAINS

Martina Vázquez, Magela Laviña, Eloísa Poey

Sección de Fisiología & Genética Bacterianas, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Tn*7* is a transposon of ca. 14 kb found in members of Gammaproteobacteria. Besides containing multiple genes devoted to transposition, it carries a class 2 integron (Int2), which typically encodes resistance to the antibiotic trimethoprim. As a transposon, it jumps by the action of a transposase, fueled by a protein with ATPase activity. An additional protein is involved in Tn*7* transposition into the chromosome, in a specific site called *attTn7*, and an alternative protein would direct it to transpose to plasmids, mainly conjugative. Thus, Tn*7* would be able to jump within cells from the chromosome to plasmids and vice versa, and move between cells being carried by conjugative plasmids.
This work focused on a collection of 18 uropathogenic *Escherichia coli* (UPEC) strains that were positive for the Int2 integrase gene (*intI2*) and, therefore, good candidates for harboring the Tn*7* transposon. To confirm its presence and to localize it at its chromosomal *attTn7* site on the chromosome, the two chromosome-Tn*7* junctions were PCR-targeted. Then, the ability of the Tn*7*-carrying replicons to horizontally transfer the transposon to an *E. coli* K12 receptor strain was tested by means of standard conjugative experiments.
Of the 18 UPEC strains, 15 contained a Tn*7* inserted in its chromosomal attachment site *attTn7*, and the remaining three, among others, resulted positive for the *intI2* gene in the plasmid fraction. Seventeen conjugation experiments UPEC x *E. coli* K12 were carried out and trimethoprim resistant clones were selected. Transfer occurred in four cases. In three of them, Tn*7* was not present on the chromosome of the UPEC strains nor in that of the transconjugants; in a second mating between these *E. coli* K12 Tn*7*$^{+}$ clones and a new *E. coli* K12 recipient strain, transconjugants grew and also resulted negative for the Tn*7* insertion on the chromosome. This indicates that in these strains Tn*7* would only be located on a conjugative plasmid and would not be able to jump to the chromosome. The fourth case of horizontal transfer of Tn*7* resulted particularly informative. The donor UPEC had a Tn*7* in the chromosome and also in the plasmid fraction. All the transconjugants analyzed resulting from the first mating received a Tn*7* that inserted in the *attTn7* site of *E. coli* K12 and, in a second mating, the same result appeared. Therefore, Tn*7* transposition into the chromosome invariably occurred when the element entered a new host, presumably being carried by a conjugative plasmid. In our collection, this would be the best example of the successful Tn*7* spread in a bacterial population, resulting from the sum of conjugation and transposition events.

A NOVEL LARGE-SCALE EPIGENETIC ANALYSIS METHOD TO UNCOVER THE CIS-REGULATORY POTENTIAL OF INDIVIDUAL TEs

Diana Voicu, Haskan Kaya, Xin Wen Yap, Michael Imbeault

University of Cambridge, Department of Genetics, Cambridge, United Kingdom

Transposable elements form the basis of approximately 50% of the human genome. There is growing evidence that some can act as cis-regulatory elements as they are a rich source of transcription factor binding sites. Moreover, the conservation of some of these elements across multiple branches of evolution speaks to their domestication potential and involvement in complex gene regulatory networks.

To understand the finer regulation of transposable elements, we analysed bulk ATAC-seq data from 93 tissues from the ENCODE repository, as well as large publicly available scATAC-seq datasets from human adult and fetal tissues. We built EpiDRAW (Epigenetic Dimensionality Reduction Analysis Workflow), a novel multi-step method which facilitates the analysis of genome-wide epigenetic datasets. It performs signal quantification, normalisation, dimensionality reduction and clustering of an arbitrarily large number of samples, and outputs tabular data and interactive graphical reports that can be used to further explore variability at select subsets of genomic regions.

We investigated differences in chromatin accessibility at transposable elements at the tissue/cell-type level. We find multiple clusters of cell-type specific accessibility at transposable elements, and we show enrichment of specific subfamilies of transposons. Moreover, we identified enrichment in specific transcription factor binding sites in these transposons, which is indicative of the potential role of these elements in gene regulatory networks. Finally, we propose a mechanism of context-dependent regulation of transposable elements by studying the relationship with KRAB zinc finger proteins, which are known to target mobile elements and promote the formation of heterochromatin.

SMALL RNAs FROM SELF-INITIATED ALLELE INDUCE RNA-DIRECTED DNA METHYLATION ON ACTIVE *AC* TRANSPOSONS IN MAIZE

Dafang Wang, Zuhair Ishraq, Michelle Scelfo, Tejbir Singh, Ryan Singh

Hofstra University, Biology, Hempstead, NY

Transposable elements (TEs) are ubiquitous genetic elements that can replicate themselves, insert into new locations in genomes, and cause mutations or large-scale chromosomal rearrangements. Small RNAs synthesized from an epigenetic pathway can effectively target and silence transposons. Here, we utilize *Ac killer* (*Ack*) in maize, a naturally occurring silencer from *Ac* alternative transposition, to investigate the initiation of epigenetic silencing of *Ac* transposons. *Ack* is a trans-acting silencer that produces the classes of 21/22 nt and 24 nt small RNAs with sequence homology to *Ac* elements and thus can target active *Ac*. We tested three growth stages in maize: germinating embryo, V1 leaf 3, and V8 leaf 10. All growth stages show significant enrichment in DNA methylation in all cytosine contexts, demonstrating RNA-dependent DNA methylation.

SINGLE-CELL RNA AND QUANTITATIVE IMMUNOFLUORESCENCE PROFILES REVEAL STRONG INTERACTIONS BETWEEN LINE-1 RETROELEMENT AND MACROPHAGES IN NON-SMALL CELL LUNG CANCERS

Yingshan Wang[1], Reid Powell[2], Kenneth S Ramos[1]

[1]Texas A&M Institute of Biosciences and Technology, Center for Genomic and Precision Medicine, Houston, TX, [2]Texas A&M Institute of Biosciences and Technology, Center for Translational Cancer Research, Houston, TX

Long-interspersed element-1 (LINE-1) is a family of active autonomous retrotransposons in the human genome that function as insertion mutagens and modulators of chromatin architecture. Active cycles of LINE-1 retrotransposon activity are recognized as distinctive features of several cancers and as key contributors to tumor heterogeneity and immune evasion. Notably, LINE-1 is negatively correlated with the expression of immunologic response genes and increased in tumors with low immune activity. In the present study we examined cell-specific profiles of LINE-1 expression across different cancers and correlated these findings with patterns of immune cell infiltration into non-small cell lung cancer (NSCLC) tissues. Single-cell RNA expression of the LINE-1 type transposase domain containing 1 (L1TD1) across 30 different tissue types was evaluated using the Human Protein Atlas database. L1TD1 was highly expressed in placenta, testis, colon, appendix, and lung. Single-cell analysis showed that L1TD1 was highly expressed in macrophages of lung tissue, but not in other tissues, and highly correlated with several macrophage marker genes, including CD68, MARCO, MRC1 and MSR1. Quantitative immunofluorescence analysis of 26 NSCLC tissues, along with matched adjacent non-tumor counterparts, confirmed in silico findings and showed a strong immunofluorescence signal that colocalized LINE-1 ORF1p with the expression of CD68, a macrophage marker, in both lung squamous cell carcinomas and lung adenocarcinomas, and to a lesser extent in non-tumor adjacent tissues. Together, these findings suggest that LINE-1 may couple intricate genetic interactions involved in the regulation of macrophage infiltration into the tumor microenvironment and highlight the importance of LINE-1 in lung oncogenesis.

POSTTRANSCRIPTIONAL REGULATION OF LINE-1 RETROTRANSPOSITION BY PROTEINS SHAPING L1 mRNA ENDS

Zbigniew Warkocki, Natalia Szostak, Arkadiusz Kajdasz, Anna Philips

Polish Academy of Sciences, Institute of Bioorganic Chemistry, Poznan, Poland

LINE-1 (L1) retrotransposons are mobile genetic elements that create new genomic insertions by a copy-paste mechanism involving L1 RNA/RNP intermediates. To do so L1 exploits target –primed reverse transcription (TPRT) wherein the poly(A) tail of L1 mRNA serves as a primer for reverse transcription by the retrotransposon ORF2p protein. We have previously demonstrated that L1 3’ ends are pervasively uridylated and regulated by multiple factors, like TUTases or XRN1, involved in shaping 3’ ends. We further explored the impact of proteins shaping mRNA 3’ ends on L1 by testing the effects of temporal depletions of PABPC and components of the deadenylase complex. We observed that depletion of either of the proteins reduces L1 retrotransposition frequency. This occurs despite the steady-state and kinetic levels of L1 mRNA and proteins are stable. However, under the depletion conditions the L1 mRNA 3’ ends change so that they are less deadenylated and, despite this, increasingly uridylated. Using dedicated L1 reporters with predefined 3' end of different poly(A) tails' lengths and uridylation we confirm that uridylation of even long poly(A)-tailed L1 reporter mRNAs reduces its ability to retropose. To test whether this might also affect ORF2p binding to L1 mRNA 3’ ends we performed RNA-IPs of ORF2p and demonstrated that it can bind uridylated L1 mRNA as effectively as polyadenylated ones. We also explore a possibility of premature cytoplasmic RT priming by ORF2p on uridylated L1 using reporters. Finally, we show that L1 reporters created with an internal poly(A) tract instead of a poly(A) tail do not retropose efficiently, and so that L1 requires poly(A) at its very 3’ end. In sum, we provide a comprehensive view on the role of L1 3’ ends in its biology and deepen the understanding of the fate of the majority of L1 mRNA/RNPs acculmulating in the cell.

Funding: Narodowe Centrum Nauki [UMO-2019/33/B/NZ1/02260] to Z.W.

MOV10 REGULATES CODING AND ncRNA METABOLISM THROUGH A URIDYLATION-MEDIATED MECHANISM

Lidia Lipinska-Zubrycka[1], Michal Malecki[1], Anna Philips[2], Zbigniew Warkocki[2]

[1]Polish Academy of Sciences, Institute of Biochemistry and Biophysics, Warsaw, Poland, [2]Polish Academy of Sciences, Institute of Bioorganic Chemistry, Poznan, Poland

MOV10 is an RNA helicase and RNPase involved in non-sense mediated decay [1,2], regulation of retrotransposons and retroviruses [3,4]. Earier works reported the importance of RNA 3' uridylation by TUTases in mRNA and ncRNA metabolism [5,6,7,8]. We created a series of knock-out MOV10, TUT4, TUT7, TUT4/7, and DIS3L2 clonal cell lines and used them to investigate a possible functional cooperation between MOV10 and the other factors in regulating cellular RNAs. By using high-throughput transcriptomics we demonstrate highly correlative gene expression changes between MOV10, TUT4/7 and DIS3L2 knock-outs in regards to both mRNA and ncRNAs. We further observe decreased TUT4/7-dependent uridylation of mRNA and ncRNA in MOV10-deficient cells and show importance of the regulatory axis involving the proteins in apoptosis. Furthermore we observed enrichement of TUT4, but not TUT7, in P-bodies and identified the region of the protein important for this localization. Interestingly, an RNA pol III-transcribed ncRNA of retrotransposonal origin snaR-A, undergoes uridylation by only one of the two TUTases, hence demonstrating some substrate selectivity of these enymes. In sum, we demonstrate the role of MOV10 in regulating pol III ncRNA and also mRNAs in a RNA decay mechanism involving TUTases.

1 Gregersen, et al. (2014) Mol. Cell, 54(4), pp. 573–585. PMID:24726324.
2 Kurosaki, et al. (2018) Nat. Struct. Mol. Biol. 25(10), pp. 940–950. PMID:30275517.
3 Warkocki, et al. (2018) Cell, 174(6), pp. 1537-1548.e29. PMID:30122351.
4 Le Pen, et al. (2018) Nat. Struct. Mol. Biol., 25(9), pp. 778–786. PMID:30104661.
5 Lim, et al. (2014) Cell, 159(6), pp. 1365–1376. PMID:25480299.
6 Łabno, et al. (2016) Nucleic Acids Res., 44(21), p. gkw649. PMID:27431325.
7 Pirouz, et al. (2016) Cell Rep., 16(7), pp. 1861–1873. PMID:27498873.
8 Warkocki, et al. (2018) Philos. Trans. R. Soc. B Biol. Sci., 373(1762), p. 20180162. PMID:30397099.

Funding: Narodowe Centrum Nauki [UMO-2019/33/B/NZ1/02260] [UMO-2017/26/D/NZ1/00887]

DISTINCTIVE REGULATION OF TRANSPOSABLE ELEMENTS IN THE HUMAN BRAIN

Risa Watanabe[1,2], Cyril J Peter[1,3], Bicheng Jiang[1,3], Foramben Patel[4], Travis Dawson[1,3], Aman Agarwal[1,3], Will Lounsbery-Scaife[2], Maya Fridrihk[5], Kiran Girdhar[6], Panos Roussos[1,2,4,6], Ashok Patowary[9], Sathiji K Nageshwaran[9], Nadejda Tsankova[3,10], Hyejung Won[11], Robert P Sebra[5], Benjamin K Chen[4], Dan Hasson[2,12], Schahram Akbarian[1,3,5]

[1]Icahn School of Medicine at Mount Sinai, Department of Psychiatry, New York, NY, [2]Icahn School of Medicine at Mount Sinai, Tisch Cancer Institute Bioinformatics for Next Generation Sequencing (BiNGS) core, New York, NY, [3]Icahn School of Medicine at Mount Sinai, Department of Neuroscience, New York, NY, [4]Icahn School of Medicine at Mount Sinai, Department of Medicine, Division of Infectious Diseases, Immunology Institute, New York, NY, [5]Icahn School of Medicine at Mount Sinai, Department of Genetics and Genomic Sciences, Center for Advanced Genomics Technology, New York, NY, [6]Icahn School of Medicine at Mount Sinai, Center for Disease Neurogenomics, New York, NY, [7]James J. Peters VA Medical Center, Mental Illness Research Education and Clinical Center (VISN 2 South), Bronx, NY, [8]James J. Peters VA Medical Center, Center for Precision Medicine and Translational Therapeutics, Bronx, NY, [9]University of California Los Angeles, Department of Neurology, Los Angeles, CA, [10]Icahn School of Medicine at Mount Sinai, Department of Pathology, Molecular, and Cell-Based Medicine, New York, NY, [11]University of North Carolina, Department of Genetics, Chapel Hill, NC, [12]Icahn School of Medicine at Mount Sinai, Department of Oncological Sciences, New York, NY

Multilayered mechanisms regulating gene expression have evolved to orchestrate the complexities of the human brain. Among these mechanisms, Transposable Elements (TEs) stand out as crucial players in shaping the genomic landscape and contributing to the diverse gene regulation. However, due to its repetitive nature, studying the loci-specific epigenetic profile of TEs have been difficult. Here, we applied fiber-seq, an adenine-methylation based long-read single-molecule chromatin accessibility assay, to sorted cerebellum nuclei and the homogenate cortex samples. This allowed us to uniquely map the epigenetic architecture, including nucleosome positioning and DNA methylation, at base-resolution along ~10kb chromatin fibers in the brain. Interestingly, TEs showed tissue-, cell type-, loci-, and family-specific chromatin accessibility patterns. To examine the relationship between TE accessibility and the chromatin structure, we examined Hi-C sequencing data from the sorted cortex nuclei. Our analyses demonstrated the TE's contribution to the genomic structure in the human brain. Supported by NIDA DP1 DA056018.

GENOME REMODELING MECHANISMS OF INSECT PIGGYBAC AND HUMAN PIGGYBAC TRANSPOSABLE ELEMENT DERIVED 5 (PGBD5) USING LONG-READ ONT SEQUENCING

Makiko Yamada[1,2], Seongmin Choi[3], Stephanie Chrysanthou[4], Juan Li[4], Neeman Mohibullah[4], Eliyahu Havasov[3], Andrew Mcpherson[3], Sohrab Shah[3], Yves Bigot[5], Alex Kentsis[1,2,6]

[1]Tow Center for Developmental Oncology, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, [2]Molecular Pharmacology Program, Sloan Kettering Institute, New York, NY, [3]Halvorsen Center for Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, [4]Integrated Genomics Operation, Memorial Sloan Kettering Cancer Center, New York, NY, [5]UMR INRAE 0085, Centre INRAe Val de Loire, Nouzilly, France, [6]Departments of Pediatrics, Pharmacology, and Physiology & Biophysics, Weill Medical College of Cornell University, New York, NY

PiggyBac Transposable Element Derived 5 (PGBD5) is a highly conserved vertebrate protein originating from a piggyBac transposase gene. It is implicated in mammalian brain development and childhood cancer pathogenesis. Although PGBD5 contains a variant catalytic triad in its RNase H-like domain, its integration mechanisms of piggyBac transposons into human chromosomes remain unknown. Indeed, several recent studies reported inability to detect PGBD5-mediated transposition. Here, we sought to define precise mechanisms associated with PGBD5 activity. First, we established optimized conditions required for the detection of PGBD5-dependent genomic integration to overcome the cytotoxicity of PGBD5 in human cells and confirmed that PGBD5 can mediate precise piggyBac transposon excisions, but with efficiency lower than that of PB transposase. To define the genomic architecture of PGBD5-induced DNA rearrangements, including those displaying canonical features of DNA transposition, we used two approaches. We used PCR assays to confirm that canonical transposition junctions so far obtained by FLEA- and LAM-PCR were not artifacts and resulted from transposon integrations due to piling of post-integration recombination events. Using whole-genome and hybridization-based targeted long-read Oxford Nanopore Technologies (ONT) sequencing, we found that PGBD5 induces DNA rearrangements, some of which involve breakpoints near piggyBac terminal repeats. These studies also identified off-target mechanisms of insect piggyBac transposition, providing one of the most complete studies to date of insect piggyBac genomic activity in human cells. We discuss these findings with respect to the physiologic and pathophysiological functions of PGBD5 in brain and cancer development, as well as the use of heterologous transposases such as piggyBac for gene transfer and gene therapy in human cells. Lastly, this study provides valuable methods for the unbiased analysis of structural variations in cells using long-read ONT sequencing.

PANGENOME ASSEMBLIES REVEAL THE EVOLUTION AND RECENT *IN VIVO* ACTIVITY OF HUMAN LINE-1 RETROTRANSPOSONS

Lei Yang[1], Amanda Norseen[1], Rick McLaughlin[1,2]

[1]Pacific Northwest Research Institute, PNRI, Seattle, WA, [2]University of Washington, Molecular & Cellular Biology, Seattle, WA

As the most abundant and only active autonomous transposable elements in extant human genomes, LINE-1s (Long INterspersed Element-1s) affect human physiology and disease risk through their ability to interact with and reorganize the genome. Because LINE-1s are long and highly repetitive, our understanding of LINE-1 variation and recent evolution within diverse humans has been limited by the scarcity of haploid resolved and long read based assemblies. Here, we provide the first analysis of LINE-1 insertions, sequence variation, and recent activity using diverse, long read resolved human genomes from the HPRC (Human Pangenome Reference Consortium). Within the 94 HPRC haplotypes, we found ~13,000 LINE-1s with both intact open reading frames required for retrotransposition located at 683 unique insertion sites. We found tremendous variation among LINE-1s from different haploid assemblies at the same insertion site, both in their sequences and potential activity. We constructed networks of these 683 LINE-1s, showing the evolutionary transitions from ancient, active groups of LINE-1s to many younger, currently active LINE-1s in modern humans. Using sequence distances computed between these 683 LINE-1s and the 3' transduction between LINE-1 insertion sites, we quantified the recent *in vivo* activity of each LINE-1, revealing the specific LINE-1s that have retrotransposed at high rates in humans. Finally, our evolutionary analysis shows the emergence of new adaptive changes which we test *in vitro* to identify putative drivers of LINE-1 evolution in humans. Our data serve as the nearly complete reference of the diversity, activity and adaptation of the youngest full-length LINE-1s in the global human population, providing novel insights into how LINE-1s affect the human genome, physiology and disease risk.

SAT-TEs: A HIDDEN SOURCE OF HUMAN GENOMIC AND REGULATORY VARIATION.

Panpan Zhang, John McCormick, Cedric Feschotte

Cornell University, MBG, Ithaca, NY

Anecdotal reports have shown that TEs can form tandem arrays of satellite DNA (which we refer to as SAT-TEs) however those structures are challenging to identify, often omitted or misassembled in genome assemblies, and a systematic characterization of SAT-TEs in the human genome is lacking. Here we took advantage of the recently released T2T reference assembly of the human genome to identify SAT-TEs and investigate their origin and variation and begin to unravel their functional impact. We identified 2120 SAT-TE loci in the T2T assembly, 10.4% of the total sequence length were unrecognizable in the previous reference genome (hg38). Certain TE subfamilies are enriched for SAT-TEs, including 47 LINE1, 8 Alu, and 44 ERV subfamilies. Most of these subfamilies also show the greatest expandability as measured by the number of tandem units per locus and they tend to originate from young TE families. The longest SAT-TE is a LTR12B locus expanded to 207 units covering 265 kb on the Y chromosome. We focused the rest of our analysis on the 94 most expanded SAT-TEs. To explore SAT-TE variability in the human population, we analyzed 47 pangenomes recently assembled from long-reads. We found extensive variation in tandem repeat number at several loci, and also differences between maternal and paternal contributions. Using phasing information, we could also infer haplotype variation in SAT-TEs. We observe a high level of heterozygosity in some individuals. In one extreme example, a L1HS locus was composed of 15 tandem units on the paternal chromosome covering 84 Kb, while that same locus was contracted to 0 unit on the maternal chromosome. Next, we investigated the cis-regulatory potential of SAT-TEs using ENCODE data for histone modifications, chromatin accessibility, and transcription factor binding. We found evidence that a subset of SAT-TEs show hallmarks of cis-regulatory activity. Lastly, we found that a subset of SAT-TE-derived enhancers occurs as extrachromosomal circular DNA in cancer cells, which may confer trans-regulatory potential. This study reveals an underappreciated abundance of SAT-TEs in the human genome and suggests that these loci are highly dynamic structural variants with likely functional impact in human evolution, disease and development.

HIV ACCESSORY PROTEIN VPR ACTIVATES TRANSPOSABLE ELEMENTS DERIVED ENHANCERS BY TARGETING HOST KRAB ZINC-FINGER

Weiyu Zhang, Manvendra Singh, Sabrina Leddy, Cedric Feschotte

Cornell University, Molecular Biology and Genetics, Ithaca, NY

Viruses are known to hijack host cellular pathways to favor their life cycle. Understanding how viruses target host cellular components can yield crucial insights into viral pathogenesis and cell biology. Our analysis of public mass spectrometry data suggests that HIV infection leads to the rapid degradation of Krüppel-associated box (KRAB) domain-containing zinc-finger proteins (KZFPs). We identify the HIV-encoded accessory protein Vpr as a direct interactor with KZFPs. The Vpr Q65R mutant, which abrogates its binding to the E3 ligase complex, abolishes its ability to target KZFPs for degradation. We further experimentally validated down-regulation of KZFPs at the protein level upon Vpr overexpression. These data suggest Vpr specifically targets KZFPs for degradation through the E3 ubiquitin-proteasome pathway, which may facilitate the HIV life cycle by: 1. Transforming heterochromatin to euchromatin to facilitate HIV integration by untethering KAP1 from DNA; 2. Activating transposable element derived cis-elements to affect the immune network; 3. Removing the epigenetic barriers of HIV transcription, such as ZNF417, which has the ability to bind to the HIV primer binding site (PBS) and repress proviral transcription. In addition to ZNF417, we focused on ZNF430 and ZNF267, which bind almost exclusively to hundreds of THE1B and THE1C LTR elements, respectively. These two subfamilies of THE1 elements are pervasive binding targets for NF-κB, a major regulator of HIV transcription, and gain H3K27ac and chromatin accessibility upon TNF-α treatment, indicative of their activity as infection-inducible enhancers. Based on these observations, we hypothesize that the targeted degradation of KZFPs via Vpr affects infection and pathogenesis by preventing ZNF417 from repressing HIV transcription and by modulating immune gene networks regulated by THE1B/THE1C-derived enhancers.

ENHANCED HAT-SEQ REVEALS CLONAL SOMATIC L1 INSERTIONS IN ATAXIA TELANGIECTASIA BRAINS

Boxun Zhao[1], Yuchen Cheng[1], Khue Nguyen[1], Shayna Mallett[1], Michael A Lodato[2], Eunjung Alice Lee[1]

[1]Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, [2]Chan Medical School, University of Massachusetts, Worcester, MA

LINE-1 (L1) retrotransposons are active mobile genetic elements that can "copy-and-paste" themselves within the human genome, contributing to genetic mosaicism in the body. Recent studies have observed somatic L1 retrotransposition in multiple human tissues, such as the brain, heart, and colon. However, our knowledge of their frequency, distribution, and functional consequences remains limited.

To detect somatic L1 insertions across a wide range of clonality, we enhanced the human active transposon sequencing (HAT-seq) method by integrating primary template-directed amplification (PTA), a recent whole-genome amplification technique. This combination, named PTA HAT-seq, enables the detection of rare somatic L1 insertions present in small cell populations or even single cells. Rigorous evaluation using 1) mixed bulk gDNA with known L1 spike-in proportions, 2) sorted micro-bulk samples (200 cells), and 3) single cells demonstrated exceptional efficiency in capturing L1 insertions, achieving >90% recovery for reference L1 insertions and >80% recovery for non-reference insertions with >800-fold and >200-fold enrichment over whole-genome sequencing, respectively. Notably, our method is sensitive and reliable for detecting L1 insertions below 0.5% clonality in bulk samples and private insertions in single cells.

We applied the enhanced HAT-seq to postmortem brain samples from 12 Ataxia Telangiectasia (A-T) patients. A-T is a rare, recessive childhood disorder with progressive cerebellar degeneration and has been associated with an increased L1 copy number. We observed a higher number of somatic L1 insertions in the cerebellum compared to the prefrontal cortex (PFC). Remarkably, in one case, we discovered 14 clonal somatic L1 insertions shared across different brain regions. We experimentally resolved 10 of these insertions, confirming their TPRT hallmarks by nested PCR and amplicon sequencing. Furthermore, we developed digital droplet PCR assays to quantify their clonalities, revealing a wide spectrum from ~40% in the cerebellum to ~0.1% in the PFC. Most of these insertions occurred during early development, and we are currently mapping their mosaicism patterns across various brain regions and non-brain tissues of this donor to trace their lineage and explore potential functional impact.

In summary, we established PTA HAT-seq, a method that enables the profiling of somatic L1 insertions in cell populations as few as a single cell. As demonstrated in our A-T study, this methodology advancement, coupled with rigorous experimental validation and quantification assays, opens new avenues for characterizing the landscape and investigating the functional impact of somatic L1 insertions in human tissues and diseases.

TUMOR SUPPRESSIVE ACTIVITIES FOR POGO TRANSPOSABLE ELEMENT DERIVED WITH KRAB DOMAIN VIA RIBOSOME BIOGENESIS RESTRICTION

Zhenbo Tu[1], Mahmoud A Bassal[2], George W Bell[3], Yanzhou Zhang[4], Yi Hu[5], Liza M Quintana[6], Deeptha Gokul[1], Daniel G Tenen[2], Antoine E Karnoub[1]

[1]Beth Israel Deaconess Medical Center and Harvard Medical School, Pathology and Cancer Center, Boston, MA, [2]Harvard Stem Cell Institute, Regenerative Biology, Cambridge, MA, [3]Whitehead Institute for Biomedical Research, , Cambridge, MA, [4] Harvard Stem Cell Institute, Regenerative Biology, Cambridge, MA, [5]Beth Israel Deaconess Medical Center and Harvard Medical School, Pathology and Cancer Center, Boston, MA, [6]Harvard Stem Cell Institute, Regenerative Biology, Boston, MA, [7]Beth Israel Deaconess Medical Center and Harvard Medical School, Pathology and Cancer Center, Boston, MA

Transposable elements (TEs) are indispensable for human development, with critical functions in pluripotency and embryogenesis. TE sequences also contribute to human pathologies, especially cancer, with documented activities as cis/trans transcriptional regulators, as sources of non-coding RNAs, and as mutagens that disrupt tumor suppressors. Despite this knowledge, little is known regarding the involvement of TE- derived genes (TEGs) in tumor pathogenesis. Here, systematic analyses of TEG expression across human cancer reveal a prominent role for pogo transposable element derived with KRAB domain (POGK). We show that POGK acts as a tumor suppressor in triple-negative breast cancer (TNBC) and that it couples with the co- repressor TRIM28 to directly block the transcription of ribosomal genes RPS16 and RPS29, in turn causing widespread inhibition of ribosomal biogenesis. We report that POGK is deactivated by isoform switching in clinical TNBC, altogether revealing its previously undescribed, exapted activities in tumor growth control.

NOTES

NOTES

NOTES

NOTES

NOTES

Participant List

Dr. Shawn Ahmed
University of North Carolina
shawn@med.unc.edu

Ernesto Arias-Palomo
CIB-CSIC
earias@cib.csic.es

Franziska Ahrend
National Institutes of Health
franziska.ahrend@nih.gov

Dr. Vladimir Arinkin
University of Geneva
vladimir.arinkin@unige.ch

Dr. Anuj Ahuja
Weill Cornell medicine
ana4028@med.cornell.edu

Dr. Irina Arkhipova
Marine Biological Laboratory
iarkhipova@mbl.edu

Prof. Yasunori Aizawa
Institute of Science Tokyo
yaizawa@bio.titech.ac.jp

Mr. Shota Azuma
RIKEN
shota.azuma@riken.jp

Dr. Asya Akca Akyildiz
Brown University
asyaakca@hacettepe.edu.tr

Dr. Pierre Baduel
CNRS/ENS
pierre.baduel@ens.psl.eu

Dr. Wenfeng An
South Dakota State University
wenfeng.an@sdstate.edu

Christopher Baker
The Jackson Laboratory
christopher.baker@jax.org

Claire An
University of Pennsylvania
anhyunh@pennmedicine.upenn.edu

Mr. Thomas Balan
CNRS - Universite Paris Cite
thomas.balan@ijm.fr

Dr. Abhishek Anand
NIH
ananda5@nih.gov

Marius Baltramonaitis
Vilnius University
marius.baltramonaitis@gmc.vu.lt

Dr. Krista Angileri
Cornell University
kma99@cornell.edu

Dr. Orsolya Barabas
University of Geneva
orsolya.barabas@unige.ch

Sokratis Apostolidis
University of Pennsylvania
sokratis.apostolidis@pennmedicine.upenn.edu

Dr. Justin Becker
Dana-Farber Cancer Institute
justinsbecker@gmail.com

Dr. Matthew Bendall
Weill Cornell Medicine
dnixon@gwu.edu

Dr. Mireille Betermier
CNRS - I2BC
mireille.betermier@i2bc.paris-saclay.fr

Katelyn Boese
Cornell University
kgb54@cornell.edu

Dr. Christopher Bolt
Tessera Therapeutics
cbolt@tesseratx.com

Dr. Emilie Brasset
University Clermont Ferrand
emilie.brasset@uca.fr

Mr. Juan Bravo
University of Southern California
juanb@usc.edu

Mr. Raphael Britt
Brown University
raphael_britt@brown.edu

Dr. Melania Bruno
National Institutes of Health
melania.bruno@nih.gov

Ms. Kerry Bubb
University of Washington
kbubb@uw.edu

Ms. Mireia Bueno Merino
Institute of Plant Sciences Paris-Saclay
mireia.bueno@universite-paris-saclay.fr

Dr. Kathleen Burns
Dana-Farber Cancer Instiute
kathleenh_burns@dfci.harvard.edu

Mr. Arad Bustan
JAX
arad.bustan@jax.org

Dr. Ivana Celic
Tulane University
icelic@tulane.edu

Dr. Ethel Cesarman
Weill Cornell
ecesarm@med.cornell.edu

Laura Chacon Machado
Cornell University
lc978@cornell.edu

Dr. Yung-Heng Chang
Stony Brook School of Medicine
Yung-Heng.Chang@stonybrook.edu

Mr. Shashank Chary
Australian National University
shashank.chary@anu.edu.au

Dr. Sihem Cheloufi
University of California Riverside
cheloufi@ucr.edu

Peiwei Chen
Cornell University
pcchen@caltech.edu

Meijuan Chen
University of California, Riverside
meijuanc@ucr.edu

Dr. Wen-Chih Cheng
Dana-Farber Cancer Institute
WEN-CHIH_CHENG@DFCI.HARVARD.EDU

Mr. Jason Chobirko
Cornell University
jdc397@cornell.edu

Ana Caroline Gandara
Morgridge Institute for Research/UW Madison
paivagandara@wisc.edu

Vivek Gandhivel
Donald Danforth Plant Science Center
vgandhivel@danforthcenter.org

Raquel Garza
Lund University
raquel.garza@med.lu.se

Dr. Yejing Ge
MD Anderson Cancer Center
YGe1@mdanderson.org

Dr. Mary Gehring
Whitehead Institute for Biomedical Research
mgehring@wi.mit.edu

Dr. George Ghanim
Princeton University
gghanim@princeton.edu

Prof. Alba Guarne
McGill University
alba.guarne@mcgill.ca

Dr. Asiya Gusa
Duke University
asiya.gusa@duke.edu

Dr. Astrid Haase
National Institutes of Health
astrid.haase@nih.gov

Mr. Morgan Hamm
University of Washington
mhamm@uw.edu

Dr. Jeffrey Han
Tulane University
jhan5@tulane.edu

Mr. Hinata Hashimoto
Tokyo Institute of Technology
hashimoto.h.aj@m.titech.ac.jp

Prof. Abeer Hassanin
Texas A&M Institute of Biosciences and Technology
ahassanin@tamu.edu

Ms. Camille Heisserer
CNRS
camille.heisserer@univ-tours.fr

Dr. Alison Hickman
Molecular Biology and Biochemistry
alisonh@niddk.nih.gov

Ms. Lilly Horowitz
Jericho High School
lillyhorowitz@gmail.com

Charlotte Huang
Hofstra University
chhu0608@gmail.com

Isabel Hubbard
Stony Brook University
isabel.hubbard@stonybrook.edu

Dr. Michael Imbeault
University of Cambridge
mi339@cam.ac.uk

Zuhair Ishraq
HOFSTRA UNIVERSITY
zishraq1@pride.hofstra.edu

Dr. Atma Ivancevic
University of Colorado Boulder
atma.ivancevic@colorado.edu

Dr. Yannick Jacob
Yale University
yannick.jacob@yale.edu

Dr. Steve Jacobsen
UCLA/HHMI
uclajacobsen@gmail.com

Dr. Rudolf Jaenisch
Whitehead Institute/MIT
jaenisch@wi.mit.edu

Mr. Jeremi Jaksina
University of Cambridge
jj515@cam.ac.uk

Zhecheng Jin
University of Pennsylvania
zhecheng@sas.upenn.edu

Dr. Valentin Joly
Yale University
valentin.joly@yale.edu

Dr. Natalie Jones
Cold Spring Harbor Laboratory
njones@cshl.edu

Mr. Leandro Jorquera
Pompeu Fabra University
leandro.jorquera@upf.edu

Dr. Lea Juznic
Genentech, Inc
juznicl@gene.com

Dr. Tetsuji Kakutani
The University of Tokyo
tkakutan@nig.ac.jp

Hannah Kania
Duke University
kania.hannah@duke.edu

Dr. Vladimir Kapitonov
Marine Biological Laboratory
vkapitonov@mbl.edu

Mr. Alon Katz
NIH
alonk379@gmail.com

Dr. Na Ke
New England Biolabs
ke@neb.com

Dr. Alex Kentsis
Memorial Sloan Kettering Cancer Center
tingues@mskcc.org

Mr. Christopher King
Yale University
chris.king@yale.edu

Hanna Kodama
Brown University
hanna_kodama@brown.edu

Mr. Kazuma Kondo
Brown University
kazumat.kondo@icloud.com

Dr. Miriam Konkel
Clemson University
mkonkel@clemson.edu

Ms. Shreya Krishnan
McGill University
shreya.krishnan@mail.mcgill.ca

Ms. Sophie Lanciano
Inserm - CNRS - Cote d'Azur University
sophie.lanciano@univ-cotedazur.fr

Dr. Nelson Lau
Boston University School of Medicine
nclau@bu.edu

Dr. Frederique Le Roux
University of Montreal
frederique.le.roux@umontreal.ca

Sabrina Leddy
Cornell University
sml327@cornell.edu

Ms. Sojung LEE
The University of Hong Kong
sjlee98@connect.hku.hk

Dr. SePil Lee
National Institutes of Health (NIH)-Bethesda
lees60@nih.gov

Mr. Aaron Lemus
University of Southern California
ajlemus@usc.edu

Dr. Henry Levin
NICHD/NIH
henry_levin@nih.gov

Prof. Ping Liang
Brock University
pliang@brocku.ca

Yu-Ching Liao
Cornell University
yl2897@cornell.edu

Dr. Hoong Chuin Lim
New England Biolabs Inc. - MA
hlim@neb.com

Ms. Winni Lin
Addition Therapeutics
wlin@additiontx.com

Ms. Yajie Liu
Hong Kong Baptist University
21482101@life.hkbu.edu.hk

Dr. Alan Lorenzetti
Lieber Institute for Brain Development
alan.lorenzetti@libd.org

Dr. Zuzana Loubalova
NIH
zuzana.loubalova@nih.gov

Dr. Nikolaos Lykoskoufis
JSR LifeSciences
nlykoskoufis@jsrlifesciences.com

Dr. Jason Lynn
Cold Spring Harbor Laboratory
jlynn@cshl.edu

Anna Mackey
Duke University
mackey842@gmail.com

Mr. Alicja Macko-Podgorni
University of Agriculture
a.macko@urk.edu.pl

Dr. Victoria Madigan
MIT
vmadigan@mit.edu

Dr. Raju Mandal
National Institutes of Health
raju.mandal@nih.gov

Ms. Natasha Ann Mariano
Cornell University
nfm25@cornell.edu

Dr. Arturo Mari-Ordonez
Gregor Mendel Institute (GMI)
arturo.mari-ordonez@gmi.oeaw.ac.at

Eirene Markenscoff-Papadimitriou
Cornell University
ecm253@cornell.edu

Dr. RaviKanthReddy Marreddy
Addition Therapeutics
rmarreddy@additiontx.com

Dr. Robert Martienssen
Cold Spring Harbor Laboratory/HHMI
martiens@cshl.edu

John Martinez
University of Rochester
john_martinez1@urmc.rochester.edu

Mr. Jacob Mathai
Yale University
jacob.mathai@yale.edu

Ms. Autumn Matthews
University of Colorado Boulder
autumn.matthews@colorado.edu

John McCormick
Cornell University
jm2268@cornell.edu

Mr. Tyler McDermott
University of Connecticut
tyler.mcdermott@uconn.edu

Mr. Jeremy McIntyre
University of California, Berkeley
jeremymcintyre@berkeley.edu

Dr. Chance Meers
Columbia University
cm4064@cumc.columbia.edu

Narmin Mekawy
Stony Brook University
Narmin.Mekawy@stonybrook.edu

Dr. Barbara Mellone
University of Connecticut
barbara.mellone@uconn.edu

Rafael Mercuri
Hospital Sirio-Libanes
rmercuri@mochsl.org.br

Dr. Thiago Miller
Hospital Sirio-Libanes
tmiller@mochsl.org.br

Mr. Paolo Mita
NYU Langone Health
paolo.mita@nyulangone.org

Dr. Yorgo Modis
University of Cambridge
ymodis@mrc-lmb.cam.ac.uk

Prof. Andrew Modzelewski
University of Pennsylvania
amodz@upenn.edu

Mr. Sokviseth Moeng
Texas A&M, Institute of Biosciences and Technology
sokvisethmoeng@tamu.edu

Dr. John Moldovan
University of Michigan
jmoldova@umich.edu

Prof. Guillermo Montoya
University of Copenhagen
guillermo.montoya@cpr.ku.dk

Daniel Morales
Florida International University
dmora127@fiu.edu

Dr. John Moran
University of Michigan Medical School
moranj@umich.edu

Santiago Morell
University of Cambridge
sm2322@cam.ac.uk

Mr. Apostolos Mourtzinos
Brown University
apostolismourtz@gmail.com

Mr. Andrea Movilli
Max Planck for Biology
movilliandrea@tuebingen.mpg.de

Dr. Heloise Muller
University at Albany
hmuller@albany.edu

Dr. Swetha Murthygowda
Stony Brook University
swetha.murthygowda@stonybrookmedicine.edu

Sultan Mussakhan
Brock University
sm20if@brocku.ca

Mr. Vamshidhar Nallamala
University of Pennsylvania
vnall@upenn.edu

Ms. Raj Nandani
South Dakota State University
Raj.Nandani@sdstate.edu

Dr. RK Narayanan
Cold Spring Harbor Laboratory
narayan@cshl.edu

Dr. Maria Nilsson Janke
Senckenberg Gesellschaft für Naturforschung
maria.nilsson-janke@senckenberg.de

Mr. Kei Nishimori
RIKEN
kei.nishimori@riken.jp

Dr. Doug Nixon
The Feinstein Institutes for Medical Research/Northwell Health
dnixon1@northwell.edu

Dr. Magnus Nordborg
Gregor Mendel Institute
magnus.nordborg@gmi.oeaw.ac.at

Dr. Shumaila Noreen
Hazara University Mansehra, Pakistan
shumailanoreen@hu.edu.pk

Dr. Faisal Nouroz
Hazara University Mansehra, Pakistan
faisalnouroz@gmail.com

Dr. Shoko Oda
The University of Tokyo
shk-etnl-209@g.ecc.u-tokyo.ac.jp

Mr. Ryuto Ohara
RIKEN IMS
ryuto.ohara@riken.jp

Dr. Ozgur Oksuz
ROME Therapeutics
ooksuz@rometx.com

Dr. Marlies Oomen
Helmholtz Munich
marlies.oomen@helmholtz-munich.de

Ms. Andrea Ordonez
Biofrontiers/CU Boulder
anor9792@colorado.edu

Prof. Valerio Orlando
King Abdullah University of Science and Technology
valerio.orlando@kaust.edu.sa

Dr. Meng Ouyang
Cold Spring Harbor Laboratory
ouyang@cshl.edu

Tyler Owens
Brown University
tyler_owens@brown.edu

Mr. Joan Parets Oliver
Brown University
joan_parets@brown.edu

Mr. Matthew Peacey
Cold Spring Harbor Laboratory
mpeacey@cshl.edu

Prof. Joseph Peters
Cornell University
joe.peters@cornell.edu

Cheng Fei Phung
NA
phungchengfei@gmail.com

Dr. Maria Poey
Facultad de Ciencias, Universidad de la Republica
mepoey@fcien.edu.uy

Dr. Leandro Quadrana
Musée d'Orsay
leandro.quadrana@bio.ens.psl.eu

Dr. Irma Querques
Max Perutz Labs Vienna
irma.querques@maxperutzlabs.ac.at

Ms. Monika Rak
Louisiana State University Health Sciences Center
mrak1@lsuhsc.edu

Dr. Kenneth Ramos
Texas A&M Institute of Biosciences and Technology
kramos@tamu.edu

Dr. Helena Reyes-Gopar
Feinstein Institutes for Medical Research
dougsanfrancisco@yahoo.com

Dr. Phoebe Rice
University of Chicago
price@uchicago.edu

Dr. Carlos Alberto Rojas Cordova
University of Geneva
Carlos.RojasCordova@unige.ch

Mr. Sourav Roy
Yale University
sourav.roy@yale.edu

Natalia Rubanova
Boston Children's Hospital
nrubanova@gmail.com

Ms. Angela Russo
LMU Munich
Angela.Russo@bmc.med.lmu.de

Ms. Karabi Saha
South Dakota State University
Karabi.Saha@sdstate.edu

Ms. Eriko Sasaki
Kyushu University
sasaki.eriko.997@m.kyushu-u.ac.jp

Prof. Daniel Sauter
University Hospital Tubingen
daniel.sauter@med.uni-tuebingen.de

Erica Sawyer
Cornell University
es986@cornell.edu

Dr. Hidetoshi Saze
Okinawa Institute of Science and Technology
hidetoshi.saze@oist.jp

Mr. Almoro Scarpa
University of Veterinary Medicine, Vienna
almorscarpa@gmail.com

Richard Schargel
Cornell University
rds278@cornell.edu

Prof. David Schatz
Yale University
david.schatz@yale.edu

Dr. Sharon Schlesinger
The Hebrew university
sharon.shle@mail.huji.ac.il

Dr. William Schneider
Rockefeller University
wschneider@rockefeller.edu

Dr. Andrea Schorn
Cold Spring Harbor Laboratory
aschorn@cshl.edu

Sierra Seifert
Duke University
sierra.seifert@duke.edu

Dr. Anna Dorothea Senft
NIH
anna.senft@nih.gov

Helene Sertznig
Cold Spring Harbor Laboratory
sertznig@cshl.edu

Ms. Samantha Shaltz
Duke University
samantha.sliwa@duke.edu

Prof. Meir Shamay
Bar-Ilan University
meir.shamay@biu.ac.il

Mr. Wei Shen
FUdan University
11111010083@fudan.edu.cn

Dr. Jason Sheperd
University of Utah
Jason.Shepherd@neuro.utah.edu

Dr. Roger Sher
Stony Brook University
roger.sher@stonybrook.edu

Dr. Virginijus Siksnys
Institute of Biotechnology Vilnius University
siksnys@ibt.lt

Dr. Joe Simorowski
Cold Spring Harbor Laboratory / HHMI
simorows@cshl.edu

Marina Smiley
New England Biolabs
msmiley@neb.com

Dr. Arian Smit
Institute for Systems Biology
asmit@systemsbiology.org

Ms. Grace Smith
University Of Miami
gxs2111@miami.edu

Prof. Chengyi Song
Yangzhou University
CYSONG@YZU.EDU.CN

Mr. Satyam Srivastav
Cornell University
sps257@cornell.edu

Jared Steranka
Rome Therapeutics
jsteranka@rometx.com

Ms. Svitlana Sushko
Max-Planck Institute for Biology
svitlana.sushko@tuebingen.mpg.de

Dr. Hillary Sussman
Genome Research, Executive Editor
hsussman@cshl.edu

Mary Talley
Cleveland Clinic
talleym3@ccf.org

Shunqing Tan
Tsinghua University
tansq21@mails.tsinghua.edu.cn

Dr. Jianli Tao
Harvard Medical School
jianli.tao@childrens.harvard.edu

Dr. Martin Taylor
Brown University
martin_taylor@brown.edu

Dr. Aristeidis Telonis
University of Miami
atelonis@med.miami.edu

Mr. Francisco Tenjo-Castano
University of Copenhagen
francisco.castano@cpr.ku.dk

Dr. Laure Teysset
Sorbonne Universite, CNRS
laure.teysset@upmc.fr

Jordan Thesier
Cornell University
jt698@cornell.edu

Beth Thompson
University of Cambridge
et528@cam.ac.uk

Mr. Sou Tomimoto
Kyushu University
sou.tomimoto@gmail.com

Dr. Maria-Elena Torres-Padilla
Institute of Epigenetics & Stem Cells
torres-padilla@helmholtz-muenchen.de

Dr. Christoph Treiber
University of Oxford
christoph.treiber@biology.ox.ac.uk

Prof. Didier Trono
EPFL
didier.trono@epfl.ch

Dr. Takuya Tsubota
National Agriculture and Food Research Organizatio
tsubota@affrc.go.jp

Prof. Jose Tubio
Universidade de Santiago
JMCTUBIO@GMAIL.COM

Ms. Heather Upton
Addition Therapeutics, Inc
uptonh@gmail.com

Dr. Trevor van Eeuwen
Rockefeller University
tvaneeuwen@rockefeller.edu

Briana Van Treeck
University of California at Berkeley
briana.van.treeck@berkeley.edu

Dr. Josien van Wolfswinkel
Yale University
josien.van.wolfswinkel@yale.edu

Ms. Martina Vazquez
Facultad de Ciencias, Universidad de la Republica
martina.vazquez@fcien.edu.uy

Ms. Diana Voicu
University of Cambridge
idv20@cam.ac.uk

Dafang Wang
HOFSTRA UNIVERSITY
dafang.wang@hofstra.edu

Yingshan Wang
Texas A&M Institute of Biosciences and Technology
yingshanwang@tamu.edu

Dr. Yunhao Wang
University of Michigan
yunhwang@med.umich.edu

Ms. Shukun Wang
Purdue University
wang4207@purdue.edu

Dr. Jianhua Wang
Columbia University
jw4562@cumc.columbia.edu

Dr. Zbigniew Warkocki
Polish Academy of Sciences
zwarkoc@gmail.com

Ms. Risa Watanabe
Icahn School of Medicine at Mount Sinai
risa.watanabe@mssm.edu

Dr. Dawn Watkins-Chow
NIH
dchow@mail.nih.gov

Kyle Watters
Arbor Biotechnologies
kwatters@arbor.bio

Prof. Alexandra Whiteley
University of Colorado, Boulder
alexandra.whiteley@colorado.edu

Dr. Makiko Yamada
Memorial Sloan Kettering Cancer Center
yamadam@mskcc.org

Dr. Lei Yang
Pacific Northwest Research Institute
leiyangly@gmail.com

Dr. Irina Yushenova
Marine Biological Laboratory
iyushenova@mbl.edu

Ms. Rimante Zedaveinyte
Columbia University
rz2610@cumc.columbia.edu

Dr. Panpan Zhang
Cornell University
pz227@cornell.edu

Mr. Tao Zhang
Tsinghua University
zhangt19@mails.tsinghua.edu.cn

Dr. Weiyu Zhang
Cornell University
wz452@cornell.edu

Prof. Zhao Zhang
Duke University
z.z@duke.edu

Dr. Yixin Zhao
Cold Spring Harbor Laboratory
yizhao@cshl.edu

Boxun Zhao
Boston Children's Hospital
boxun.zhao@childrens.harvard.edu

David Zimmerman
Cold Spring Harbor Laboratory
zimmerman@cshl.edu

Dr. Sonia Zumalave Duro
EMBL-EBI
sonia@ebi.ac.uk

CODE OF CONDUCT FOR ALL PARTICIPANTS IN CSHL MEETINGS

Cold Spring Harbor Laboratory (CSHL or the Laboratory) is dedicated to pursuing its twin missions of research and education in the biological sciences. The Laboratory is committed to fostering a working environment that encourages and supports unfettered scientific inquiry and the free and open exchange of ideas that are the hallmarks of academic freedom. To this end, the Laboratory aims to maintain a safe and respectful environment that is free from harassment and discrimination for all attendees of our meetings and courses as well as associated support staff, in accordance with federal, state and local laws.

Consistent with the Laboratory's missions, commitments and policies, the purpose of this Code is to set forth expectations for the professional conduct of all individuals participating in the Laboratory's meetings program, both in person and virtually, including organizers, session chairs, invited speakers, presenters, attendees and sponsors. This Code's prohibition against discrimination and harassment is consistent with the Laboratory's internal policies governing conduct by its own faculty, trainees, students and employees.

By registering for and attending a CSHL meeting, either in person or virtually, participants agree to:

1. Treat fellow meeting participants and CSHL staff with respect, civility and fairness, without bias based on sex, gender, gender identity or expression, sexual orientation, race, ethnicity, color, religion, nationality or national origin, citizenship status, disability status, veteran status, marital or partnership status, age, genetic information, or any other criteria prohibited under applicable federal, state or local law.

2. Use all CSHL facilities, equipment, computers, supplies and resources responsibly and appropriately if attending in person, as you would at your home institution.

3. Abide by the CSHL Meeting Alcohol Policy (*see below)*.

Similarly, meeting participants agree to refrain from:

1. Harassment and discrimination, either in person or online, in violation of Laboratory policy based on actual or perceived sex, pregnancy status, gender, gender identity or expression, sexual orientation, race, ethnicity, color, religion, creed, nationality or national origin, immigration or citizenship status, mental or physical disability status, veteran status, military status, marital or partnership status, marital or partnership status, familial status, caregiver status, age, genetic information, status as a victim of domestic violence, sexual violence, or stalking, sexual reproductive health decisions, or any other criteria prohibited under applicable federal, state or local law.

2. Sexual harassment or misconduct.

3. Disrespectful, uncivil and/or unprofessional interpersonal behavior, either in person or online, that interferes with the working and learning environment.

4. Misappropriation of Laboratory property or excessive personal use of resources, if attending in person.

BREACHES OR VIOLATIONS OF THE CODE OF CONDUCT

Cold Spring Harbor Laboratory aims to maintain in-person and virtual conference environments that accord with the principles and expectations outlined in this Code of Conduct. Meeting organizers are tasked with providing leadership during each meeting, and may be approached informally about any breach or violation. Breaches or violations should also be reported to program leadership in person or by email:

- Dr. David Stewart, Grace Auditorium Room 204, 516-367-8801 or x8801 from a campus phone, stewart@cshl.edu
- Dr. Charla Lambert, Hershey Laboratory Room 214, 516-367-5058 or x5058 from a campus phone, clambert@cshl.edu

Reports may be submitted by those who experience harassment or discrimination as well as by those who witness violations of the behavior laid out in this Code.

The Laboratory will act as needed to resolve the matter, up to and including immediate expulsion of the offending participant(s) from the meeting, dismissal from the Laboratory, and exclusion from future academic events offered by CSHL.

If you have questions or concerns, you can contact the meeting organizers, CSHL staff.

For meetings and courses funded by NIH awards:

Participants may contact the Health & Human Services Office for Civil Rights (OCR). See this page for information on filing a civil rights complaint with the OCR; filing a complaint with CSHL is not required before filing a complaint with OCR, and seeking assistance from CSHL in no way prohibits filing complaints with OCR. You may also notify NIH directly about sexual harassment, discrimination, and other forms of inappropriate conduct at NIH-supported events.

For meetings and courses funded by NSF awards:

Participants may file a complaint with the NSF. See this page for information on how to file a complaint with the NSF.

Law Enforcement Reporting:

- For on-campus incidents, reports to law enforcement can be made to the Security Department at 516-367-5555 or x5555 from a campus phone.
- For off-campus incidents, report to the local department where the incident occurred.

In an emergency, dial 911.

DEFINITIONS AND EXAMPLES

Uncivil/disrespectful behavior is not limited to but may take the following forms:

- Shouting, personal attacks or insults, throwing objects, and/or sustained disruption of talks or other meeting-related events

Harassment is any unwelcome verbal, visual, written, or physical conduct that occurs with the purpose or effect of creating an intimidating, hostile, degrading, humiliating, or offensive environment or unreasonably interferes with an individual's work performance. Harassment is not limited to but may take the following forms:

- Threatening, stalking, bullying, demeaning, coercive, or hostile acts that may have real or implied threats of physical, professional, or financial harm
- Signs, graphics, photographs, videos, gestures, jokes, pranks, epithets, slurs, or stereotypes that comment on a person's sex, gender, gender identity or expression, sexual orientation, race, ethnicity, color, religion, nationality or national origin, citizenship status, disability status, veteran status, marital or partnership status, age, genetic information, or physical appearance

Sexual Harassment includes harassment on the basis of sex, sexual orientation, self-identified or perceived sex, gender expression, gender identity, and the status of being transgender. Sexual harassment is not limited to sexual contact, touching, or expressions of a sexually suggestive nature. Sexual harassment includes all forms of gender discrimination including gender role stereotyping and treating employees differently because of their gender. *Sexual misconduct* is not limited to but may take the following forms:

- Unwelcome and uninvited attention, physical contact, or inappropriate touching
- Groping or sexual assault
- Use of sexual imagery, objects, gestures, or jokes in public spaces or presentations
- Any other verbal or physical contact of a sexual nature when such conduct creates a hostile environment, prevents an individual from fulfilling their professional responsibilities at the meeting, or is made a condition of employment or compensation either implicitly or explicitly

MEETING ALCOHOL POLICY

Consumption of alcoholic beverages is not permitted in CSHL's public areas other than at designated social events (wine and cheese reception, picnic, banquet, etc.), in the Blackford Bar, or under the supervision of a licensed CSHL bartender.

No provision of alcohol by meeting sponsors is permitted unless arranged through CSHL.

Meeting participants consuming alcohol are expected to drink only in moderation at all times during the meeting.

Excessive promotion of a drinking culture at any meeting is not acceptable or tolerated by the Laboratory. No meeting participant should feel pressured or obliged to consume alcohol at any meeting-related event or activity.

VISITOR INFORMATION

EMERGENCY (to dial outside line, press 3+1+number)	
CSHL Security	**516-367-8870 (x8870 from house phone)**
CSHL Emergency	**516-367-5555 (x5555 from house phone)**
Local Police / Fire	**911**
Poison Control	**(3) 911**

CSHL SightMD Center for Health and Wellness ***(call for appointment)*** Dolan Hall, East Wing, Room 111 **cshlwellness@northwell.edu**	**516-422-4422** **x4422 from house phone**
Emergency Room **Huntington Hospital** 270 Park Avenue, Huntington	**631-351-2000**
Dentists Dr. William Berg Dr. Robert Zeman	 **631-271-2310** **631-271-8090**
Drugs - 24 hours, 7 days Rite-Aid 391 W. Main Street, Huntington	**631-549-9400**

GENERAL INFORMATION

Meetings & Courses Main Office

Hours during meetings: M-F 9am – 9pm, Sat 8:30am – 1pm
After hours – See information on front desk counter
For assistance, call Security at 516-367-8870
(x8870 from house phone)

Dining, Bar

Blackford Dining Hall (main level):
Breakfast 7:30–9:00, Lunch 11:30–1:30, Dinner 5:30–7:00
Blackford Bar (lower level): 5:00 p.m. until late

House Phones

Grace Auditorium, upper / lower level; Cabin Complex; Blackford Hall; Dolan Hall, foyer

Books, Gifts, Snacks, Clothing

CSHL Bookstore and Gift Shop
516-367-8837 (hours posted on door)
Grace Auditorium, lower level.

Computers, E-mail, Internet access

Grace Auditorium
Upper level: E-mail and printing in the business center area
WiFi Access: GUEST (no password)

Announcements, Message Board Mail, ATM, Travel info

Grace Auditorium, lower level

Russell Fitness Center
Dolan Hall, east wing, lower level
PIN#: **(On your registration envelope)**

Laundry Machines
Dolan Hall, lower level

Photocopiers, Journals, Periodicals, Books
CSHL Main Library
Open 24 hours (with PIN# or CSHL ID)
Staff Hours: 9:00 am – 9:00 pm
Use PIN# (On your registration envelope) to enter Library
See Library staff for photocopier code.
Library room reservations (hourly) available on request between 9:00 am – 9:00 pm

Swimming, Tennis, Jogging, Hiking
June–Sept. Lifeguard on duty at the beach. 12:00 noon–6:00 p.m.
Two tennis courts open daily.

Local Interest

Fish Hatchery	631-692-6758
Sagamore Hill	516-922-4788
Whaling Museum	631-367-3418
Heckscher Museum	631-351-3250
CSHL DNA Learning Center	x 5170

New York City
Helpful tip -
Take CSHL Shuttle OR Uber/Lyft/Taxi to Syosset Train Station
Long Island Railroad to Penn Station
Train ride about one hour.

TRANSPORTATION

Limo, Taxi

Syosset Limousine	516-364-9681
Executive Limo Service	516-826-8172
Limos Long Island	516-400-3364
Syosset Taxi	516-921-2141
Orange & White Taxi	631-271-3600
Uber / Lyft	

Trains

Long Island Rail Road	718-217-LIRR (5477)
Amtrak	800-872-7245
MetroNorth	877-690-5114
New Jersey Transit	973-275-5555

CSHL Campus Map

CSHL Map