Bioinformatics Advances, 2024, **00**, vbae063
https://doi.org/10.1093/bioadv/vbae063
Advance Access Publication Date: 8 May 2024
Application Note

Databases and ontologies

rNMPID: a database for riboNucleoside MonoPhosphates in DNA

Jingcheng Yang [1,2,†], **Mo Sun** [3,†], **Zihan Ran** [4,5], **Taehwan Yang** [3], **Deepali L. Kundnani** [3], **Francesca Storici** [3,*], **Penghao Xu** [3,*]

[1] State Key Laboratory of Genetic Engineering, School of Life Sciences, Human Phenome Institute, and Shanghai Cancer Center, Fudan University, Shanghai 200438, China
[2] Greater Bay Area Institute of Precision Medicine, Guangzhou, Guangdong 511462, China
[3] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, United States
[4] Department of Research, Shanghai University of Medicine & Health Sciences Affiliated Zhoupu Hospital, Shanghai 201318, China
[5] Inspection and Quarantine Department, The College of Medical Technology, Shanghai University of Medicine & Health Sciences, Shanghai 201318, China

*Corresponding authors. School of Biological Sciences, Georgia Institute of Technology, Atlanta 30332, GA, United States. E-mails: xph6113@gmail.com (P.X.) and storici@gatech.edu (F.S.)

† These authors contributed equally

Associate Editor: Lina Ma

Abstract

Motivation: Ribonucleoside monophosphates (rNMPs) are the most abundant non-standard nucleotides embedded in genomic DNA. If the presence of rNMP in DNA cannot be controlled, it can lead to genome instability. The actual regulatory functions of rNMPs in DNA remain mainly unknown. Considering the association between rNMP embedment and various diseases and cancer, the phenomenon of rNMP embedment in DNA has become a prominent area of research in recent years.

Results: We introduce the rNMPID database, which is the first database revealing rNMP-embedment characteristics, strand bias, and preferred incorporation patterns in the genomic DNA of samples from bacterial to human cells of different genetic backgrounds. The rNMPID database uses datasets generated by different rNMP-mapping techniques. It provides the researchers with a solid foundation to explore the features of rNMP embedded in the genomic DNA of multiple sources, and their association with cellular functions, and, in future, disease. It also significantly benefits researchers in the fields of genetics and genomics who aim to integrate their studies with the rNMP-embedment data.

Availability and implementation: rNMPID is freely accessible on the web at https://www.rnmpid.org.

1 Introduction

Ribonucleoside triphosphates (rNTPs), the basic building blocks of RNA, are abundantly incorporated into DNA in the form of ribonucleoside monophosphates (rNMPs) by DNA polymerases due to their similarity to DNA nucleotides (Nick McElhinny *et al.* 2010, Brown and Suo 2011). Previous studies suggested that the incorporated rNMPs constitute the most abundant non-standard nucleotides in the DNA of eukaryotic and prokaryotic genomes (Cerritelli and Crouch 2016). Genome instability can result from failure to remove the genomic rNMPs, as the presence of rNMPs in DNA can alter the DNA structure, cause chromosomal fragility, and affect protein-DNA binding activity (Chiu *et al.*, 2014, Williams *et al.*, 2016, Klein 2017). Furthermore, mutations in genes encoding any of the subunits of ribonuclease (RNase) H2, the main enzyme that initiates the rNMP removal, are found in the genotype of many patients affected by Aicardi-Goutières Syndrome (AGS) and Systemic Lupus Erythematosus (SLE), and tumor cell in skin and colorectal cancer (Williams *et al.* 2016, Moss *et al.* 2017, Hiller *et al.* 2018, Aden *et al.* 2019). On the other hand, embedded rNMPs may also have physiological roles. For example, the abundant presence of rNMPs on the leading strand of DNA replication can guide mismatch repair in eukaryotic cells (Ghodgaonkar *et al.* 2013, Williams *et al.* 2013). Despite the threat posed to genomic integrity, abundant rNMP incorporation has persisted throughout the evolutionary scale, with millions of rNMPs in the human genome (Sassa *et al.* 2019). Therefore, there is still much to uncover about the functions and consequences of rNMPs embedded in DNA.

Over the last decade, researchers have devised multiple molecular biology techniques for mapping the genomic rNMPs, including ribose-seq (Koh *et al.* 2015), emRiboSeq (Ding *et al.* 2015), Alk-HydEn-seq (Clausen *et al.* 2015), RHII-HydEn-seq (Zhou *et al.* 2019). Ribose-seq, retains the rNMP and its upstream proximal sequence in sequenced products and relies on tRNA ligase activity which can ligate a 2’3’-cyclic phosphate to a 5’-phosphate. HydEn-seq can tag the 5’-OH thus generated, placing the rNMP at -1 position when reads are aligned to the reference sequence. Compared with

Received: December 12, 2023; **Revised:** March 27, 2024; **Editorial Decision:** April 20, 2024; **Accepted:** May 7, 2024

Alk-HydEn-seq, RHII-HydEn-seq cleaves upstream the embedded rNMPs by RNase H2 and uses restriction digestion to introduce internal standards. emRiboSeq also utilizes RNase H2 to cleave 5' of an rNMP in virtro, generating a 3'-OH subsequently ligated to a tag. In this case, the rNMP is at the +1 position relative to the tagged base (Fig. 1). Researchers have created over 300 libraries of embedded rNMPs in seven different species. These libraries help scientists understand how often rNMPs are included in DNA, the patterns of rNMP embedment, and how the rNMP presence relates to DNA metabolic functions in various organisms (Nick McElhinny *et al.* 2010, Kasiviswanathan and Copeland 2011, Balachander *et al.* 2020, El-Sayed *et al.* 2021, Xu and Storici 2021a, Xu *et al.* 2024). Additionally, a series of bioinformatics tools have been developed to facilitate the mapping and analysis of rNMP incorporation in DNA (Gombolay *et al.* 2019, Gombolay and Storici 2021, Xu and Storici 2021b, 2021c). With the help of sequencing-based rNMPs mapping technology, previous studies have utilized genome-wide rNMPs as indicators of replicative polymerases (Daigaku *et al.* 2015, Sriramachandran *et al.* 2020, Koyanagi *et al.* 2022).

Other investigations have expanded our understanding of rNMP-incorporation functions, not only by highlighting the direct interplay between ribonucleotide excision repair (RER) and topoisomerase 1 (Top1), two pathways for rNMPs removal, with the transcriptional processes in eukaryotic cells (Reijns *et al.* 2022, Hao *et al.* 2023), but also by linking features of the hmtDNA rNMPs to specific characteristics of human mitochondrial DNA (hmtDNA), including Pol γ rNMP-incorporation preference, hmtDNA replication, and transcription (Xu *et al.* 2024). The investigation into how rNMP embedment affects crucial cellular activities, including replication and transcription, along with its links to diseases, demands more research. This necessity has elevated research about rNMP embedment to a key focus in recent scientific studies. To support these efforts, it is becoming increasingly important for the scientific community to create a comprehensive, multi-sourced database of rNMPs.

2 Methods

The rNMPID database is implemented by integrating more than ten published rNMP datasets derived from various species, including *Saccharomyces cerevisiae, Saccharomyces paradoxus, Schizosaccharomyces pombe, Chlamydomonas reinhardtii, Escherichia coli, Mus musculus*, and *Homo sapiens* (Supplementary Table S1). These rNMP datasets are constructed using four different rNMP-mapping techniques, ribose-seq (Koh *et al.* 2015), emRiboSeq (Ding *et al.* 2015), Alk-HydEn-seq (Clausen *et al.* 2015), and RHII-HydEn-seq (Zhou *et al.* 2019). Collected rNMP libraries are formatted in BEDGRAPH format and then converted to BigWig files for the Genome Browser (Supplementary Fig. S1). Afterward, by calculating the frequency and composition of rNMP embedment in various genetic elements, including genes, coding sequences, non-coding RNA, and other elements, we provide the researchers with valuable tools to study the association of rNMPs with such DNA elements and their role in various DNA metabolic processes. The count of rNMPs are normalized on the total number of rNMPs in the chosen rNMP library and the total nucleotide frequency from the background reference genome as previously described (Xu and Storici 2021c).

The rNMPID database is built using Rust (Matsakis and Klock 2014), TypeScript (Bierman *et al.* 2014), PostgreSQL (Momjian 2001) and additional libraries including Tokio (Tokio Team 2023), SQLx (Launchbadge Team 2023), Reactjs (Rawat and Mahajan 2020), Plotly (Johnson *et al.* 2012), Ant Design (Ant Design Team 2023), and JBrowse (Buels *et al.* 2016), etc It comprises four different modules, namely Sample Analysis, Genome Browser, Download, and

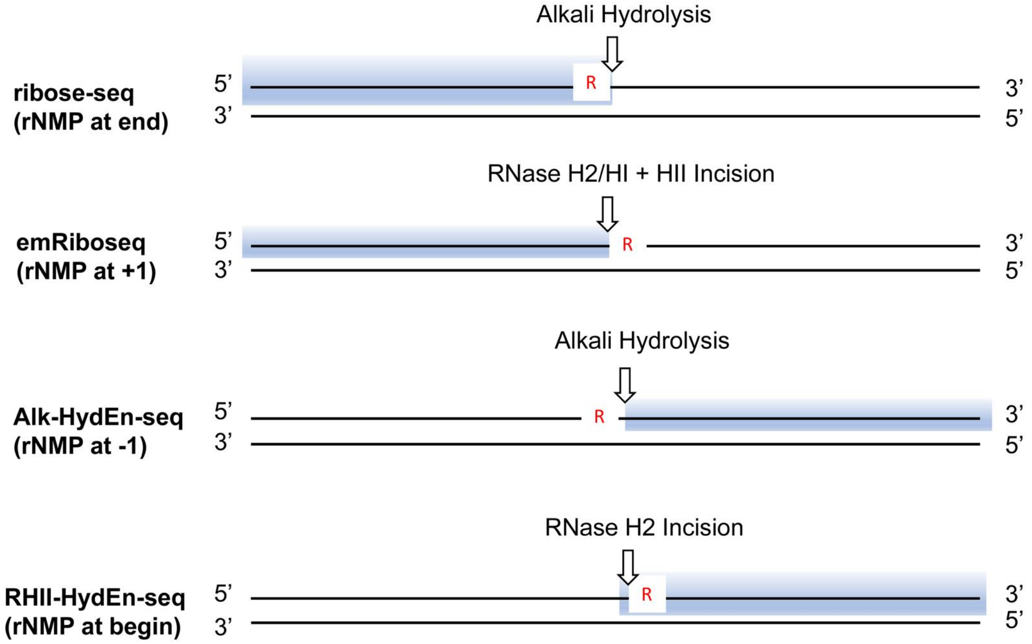


Figure 1. Different tagging strategy to map rNMPs. The position of rNMPs in DNA relative to the sites of alkaline or RNase H2 incision (arrow) and the nucleotides region captured (light blue box) in library preparation. ribose-seq directly captures rNMPs along with the nucleotides upstream from them, emRiboSeq captures the nucleotides upstream from rNMPs, Alk-HydEn-seq and RHII-HydEn-seq capture the nucleotides downstream from rNMPs, w/o or with the rNMPs, respectively.

Resource (Fig. 2A). These modules offer the opportunity to conduct a comprehensive analysis and personalized visualization of rNMP embedment in DNA.

3 Results

3.1 Sample analysis module

The Sample Analysis module provides the users with handy tools to directly analyze the rNMP-embedment samples and reveal the rNMP-embedment characteristics. The metadata table in the module contains essential information about each collected sample, including sample name, rNMP-mapping technique, species, reference genome, RNase H2 genotype, genotype, GEO accession, SRA accession, and description (Fig. 2B). Users can select multiple rNMP samples of interests to perform analyses and comparisons. Moreover, the Sample Analysis module supports multiple query methods. Users can analyze the rNMP-embedment characteristics of the whole genome or focus on a single strand, on specific DNA fragments, and chromosomes to reveal the local rNMP-embedment patterns in the selected regions (Fig. 2C). Besides, users have the ability to conveniently search for genes by directly entering the Gene ID or Gene Name into the provided search bar. This functionality enables users to swiftly retrieve rNMP features specific to their gene region of interest. By incorporating the RibosePreferenceAnalysis tool (Xu and Storici 2021c), the Sample Analysis module performs various analysis on rNMP-embedment characteristics, including bar plots showing the raw and normalized frequency of the embedded rNMPs (Fig. 2D), heatmaps showing the rNMP-embedment compositions, and heatmaps showing the preferred dinucleotide patterns of rNMP embedment (Fig. 2E). By developing the Sample Analysis module, we devised to provide the researchers with handful tools to reveal the rNMP-embedment characteristics in the selected regions of interest of different samples, which can be easily integrated into their studies with the help of customizable visualizations.

3.2 Genome browser

To show the rNMP-embedment location and frequency on a single-nucleotide level, we incorporated the JBrowse Genome Browser in our rNMPs database (Buels *et al.* 2016). The Genome Browser contains various tracks, including reference sequences, gene annotations, and rNMP location and frequency (Fig. 2F). Users can easily zoom in to their region/gene of interest to see the frequency of rNMP embedment on each nucleotide and select multiple libraries to compare.

3.3 Resources

To assist users in performing their novel analysis on rNMP samples, we provided the free download of collected rNMP sample data, reference genome, genome annotations, and formatted data showing the rNMP-embedment characteristics used in the rNMPID database. We also gathered all the useful tools and studies in the Resources modules, including rNMP-mapping techniques, bioinformatics tools for rNMP-embedment analysis, and previous rNMP-related papers.

4 Discussion and conclusion

The rNMPID database is a large-scale database. By initially integrating 326 rNMP-embedment libraries in seven different species and four different rNMP-mapping techniques, rNMPID contains 2, 789, 282, 366 unique rNMP-incorporation loci. This data amount is significantly larger than any other rNMP-related study.

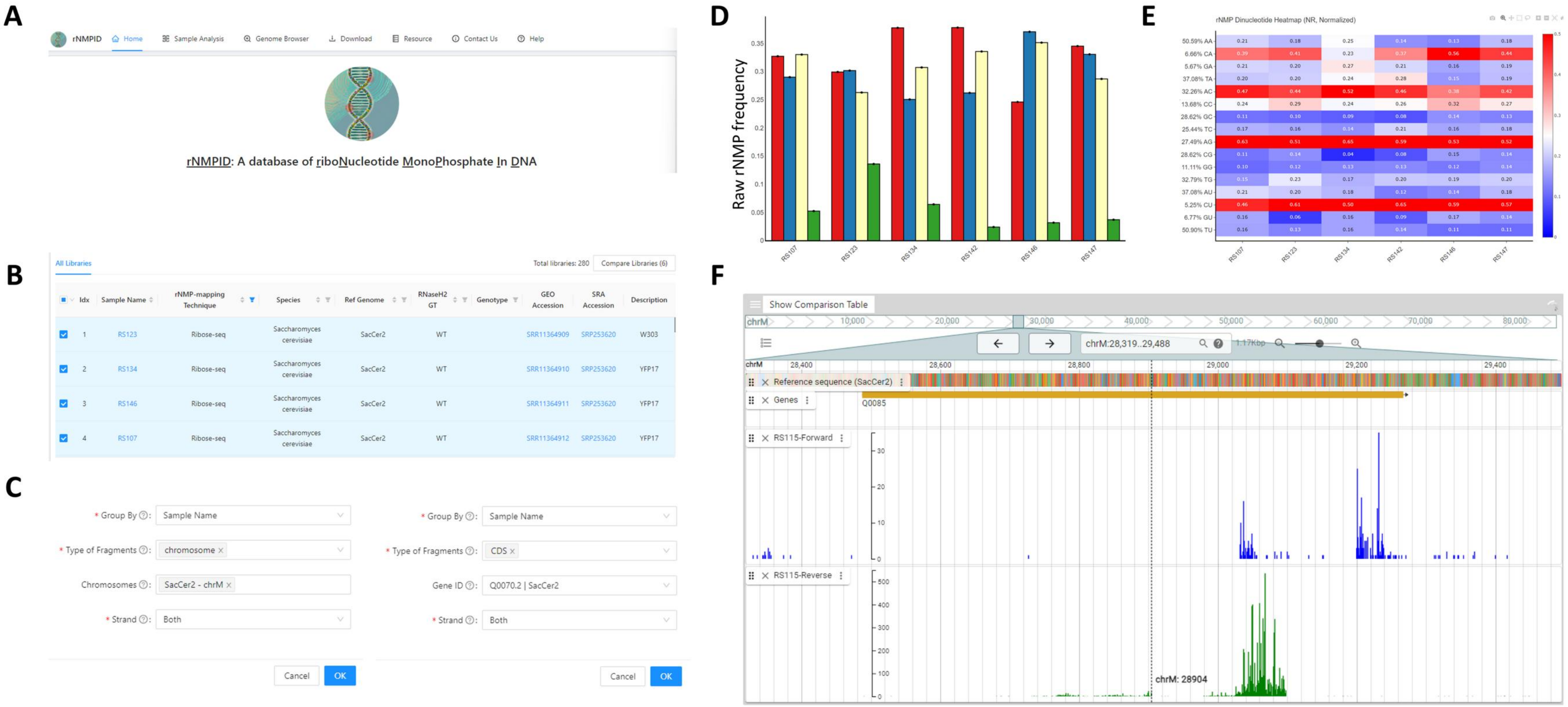
The rNMPID database provides powerful data analysis and highly customizable visualization. We have incorporated state-of-the-art tools for the rNMP-embedment analysis in the rNMPID database. These tools offer researchers a comprehensive set of five distinct visualizations dedicated to examining rNMP composition and patterns in the Sample Analysis module. Each visualization is highly customizable, allowing users to modify parameters such as sample order, scale, and grouping methods to suit their specific research requirements. Additionally, the Genome Browser module and Download module empower users to perform in-depth investigations on the genomic region of interest.

Researchers can easily reveal the distinctive rNMP-incorporation characteristics across various species, cell types, and genotypes using rNMPID. We utilized rNMPID to replicate some key findings from a previous study of rNMP-incorporation characteristics within six wild-type and eight *rnh201*-null *S. cerevisiae* libraries (Balachander *et al.* 2020). In the Sample Analysis module, we initially selected these libraries from the metatable. Subsequently, we chose "RNH2 Genotype" and "nuclear DNA" as the criteria for "Group By" and "Type of Fragments" options. This analytical approach effectively reproduced our major findings. Notably, our results revealed that rC emerged as the most abundant rNMP within the nuclear DNA of both wild-type and *rnh201*-null cells, while rU appeared as the least abundant rNMP in *rnh201*-null cells (Supplementary Fig. S2A and B). Moreover, our examination of the Genome Browser revealed the presence of short nucleotide-repeated sequences, each displaying distinct patterns of rNMP enrichment. An illustrative example can be found at locus chrM: 63583–63651 within the RS156 library, where we selected RS156 in Genome Browser modules and chrM in the comparison table. Here, rNMPs were identified at the G-nucleotide position within the TAAGTA-repeated sequence on the forward strand and at the C-nucleotide position within the TACTTA-repeated sequence on the reverse strand (Supplementary Fig. S2C).

In summary, the rNMPID database is the first database of rNMP embedment in genomic DNA, which reveals rNMP-embedment characteristics, strand bias, and preferred rNMP patterns observed in the genome of different species in rNMP libraries generated using different rNMP-mapping techniques, and in DNA samples of various genotypes. The rNMPID database provides the researchers with a solid foundation to explore the function of rNMPs embedded in genomic DNA and their association with DNA metabolic functions and potential disease.

Acknowledgements

We thank C. Gee from the Academic and Research IT for advice and technique support, and the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology for their research cyberinfrastructure resources and services. We acknowledge Y. Lee and S. Randhawa for critically reading the manuscript; and all members of the Storici Laboratory for assistance and feedback on this study. We also thank Amazon Web Services (AWS) for Promotional Credit award and device support.



All Libraries

Total libraries: 280 Compare Libraries (6)

Idx	Sample Name	rNMP-mapping Technique	Species	Ref Genome	RNaseH2 GT	Genotype	GEO Accession	SRA Accession	Description
1	RS123	Ribose-seq	Saccharomyces cerevisiae	SacCer2	WT		SRR11364909	SRP253620	W303
2	RS134	Ribose-seq	Saccharomyces cerevisiae	SacCer2	WT		SRR11364910	SRP253620	YFP17
3	RS146	Ribose-seq	Saccharomyces cerevisiae	SacCer2	WT		SRR11364911	SRP253620	YFP17
4	RS107	Ribose-seq	Saccharomyces cerevisiae	SacCer2	WT		SRR11364912	SRP253620	YFP17

* Group By ⓘ: Sample Name
* Type of Fragments ⓘ: chromosome ×
Chromosomes ⓘ: SacCer2 - chrM ×
* Strand ⓘ: Both
Cancel OK

* Group By ⓘ: Sample Name
* Type of Fragments ⓘ: CDS ×
Gene ID ⓘ: Q0070.2 | SacCer2
* Strand ⓘ: Both
Cancel OK

rNMP Dinucleotide Heatmap (NR, Normalized)

	RS107	RS123	RS134	RS142	RS146	RS147
50.59% AA	0.21	0.18	0.25	0.14	0.13	0.18
6.66% CA	0.39	0.41	0.23	0.37	0.56	0.44
5.67% GA	0.21	0.20	0.27	0.21	0.16	0.19
37.08% TA	0.20	0.20	0.24	0.28	0.15	0.19
32.26% AC	0.47	0.44	0.52	0.46	0.38	0.42
13.68% CC	0.24	0.29	0.24	0.26	0.32	0.27
28.62% GC	0.11	0.10	0.09	0.08	0.14	0.13
25.44% TC	0.17	0.16	0.14	0.21	0.16	0.18
27.49% AG	0.63	0.51	0.65	0.59	0.53	0.52
28.62% CG	0.11	0.14	0.04	0.08	0.15	0.14
11.11% GG	0.10	0.12	0.13	0.13	0.12	0.14
32.79% TG	0.15	0.23	0.17	0.20	0.19	0.20
37.08% AU	0.21	0.20	0.18	0.12	0.14	0.18
5.25% CU	0.46	0.61	0.50	0.65	0.59	0.57
6.77% GU	0.16	0.06	0.16	0.09	0.17	0.14
50.90% TU	0.16	0.13	0.16	0.14	0.11	0.11

Figure 2. rNMPID database. (A) Homepage of the rNMPID database. (B) rNMP sample metadata containing sample name, rNMP-mapping technique, species, reference genome, RNase H2 genotype, genotype, GEO accession, SRA accession, and description in Sample analysis module. (C) Users can select different type of fragment and further selection on chromosome and gene id in Sample Analysis module. (D) Example of barplot showing the composition of embedded rNMPs; red, rAMP; blue, rCMP; yellow, rGMP; green, rUMP. (E) Example of heatmap showing the normalized frequency of dinucleotides composed of the incorporated rNMP (R: rA, rC, rG, or rU) and its upstream dNMP neighbor (N: dA, dC, dG, or dT) (NR). (F) An example of Genome Browser module with reference genome sequence, gene annotations, and rNMP-embedment sites.

Supplementary data

Supplementary data are available at *Bioinformatics Advances* online.

Conflict of interest

None declared.

Funding

This work has been supported by the National Institutes of Health [NIEHS R01 ES026243 to F.S.]; Howard Hughes Medical Institute Faculty Scholars Award [HHMI 55108574 to F.S.], Mathers Foundation [AWD-002589 to F.S.]; and the W. M. Keck Foundation grants [to F.S.].

Data availability

The data underlying this article are publicly available in https://www.rnmpid.org.

References

Aden K, Bartsch K, Dahl J *et al.* Epithelial RNase H2 maintains genome integrity and prevents intestinal tumorigenesis in mice. *Gastroenterology* 2019;**156**:145–59.e19.

Ant Design Team: an enterprise-class UI design language and React UI library, 2023. https://github.com/ant-design/ant-design/ (4 April 2023, date last accessed).

Balachander S, Gombolay AL, Yang T *et al.* Ribonucleotide incorporation in yeast genomic DNA shows preference for cytosine and guanosine preceded by deoxyadenosine. *Nat Commun* 2020;**11**:2447.

Bierman G, Abadi M, Torgersen M. Understanding typescript. In: *ECOOP 2014–Object-Oriented Programming: 28th European Conference, Uppsala, Sweden, July 28–August 1 2014. Proceedings 28*. Springer Berlin Heidelberg, 2014.

Brown JA, Suo Z. Unlocking the sugar 'steric gate' of DNA polymerases. *Biochemistry* 2011;**50**:1135–42.

Buels R, Yao E, Diesh CM *et al.* JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol*2016;**17**:66–12.

Cerritelli SM, Crouch RJ. The balancing act of ribonucleotides in DNA. *Trends Biochem Sci*2016;**41**:434–45.

Chiu H-C, Koh KD, Evich M *et al.* RNA intrusions change DNA elastic properties and structure. *Nanoscale* 2014;**6**:10009–17.

Clausen AR, Lujan SA, Burkholder AB *et al.* Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat Struct Mol Biol* 2015;**22**:185–91.

Daigaku Y, Keszthelyi A, Müller CA *et al.* A global profile of replicative polymerase usage. *Nat Struct Mol Biol* 2015;**22**:192–8.

Ding J, Taylor MS, Jackson AP *et al.* Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. *Nat Protoc*2015;**10**:1433–44.

El-Sayed WMM, Gombolay AL, Xu P *et al.* Disproportionate presence of adenosine in mitochondrial and chloroplast DNA of Chlamydomonas reinhardtii. *iScience* 2021;**24**:102005.

Ghodgaonkar MM, Lazzaro F, Olivera-Pimentel M *et al.* Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol Cell* 2013;**50**:323–32.

Gombolay AL, Vannberg FO, Storici F *et al.* Ribose-Map: a bioinformatics toolkit to map ribonucleotides embedded in genomic DNA. *Nucleic Acids Res* 2019;**47**:e5- e5.

Gombolay AL, Storici F. Mapping ribonucleotides embedded in genomic DNA to single-nucleotide resolution using Ribose-Map. *Nat Protoc* 2021;**16**:3625–38.

Hao Z, Gowder M, Proshkin S *et al.* RNA polymerase drives ribonucleotide excision DNA repair in E. coli. *Cell* 2023;**186**:2425–37.e21.

Hiller B, Hoppe A, Haase C *et al.* Ribonucleotide excision repair is essential to prevent squamous cell carcinoma of the skin. *Cancer Res* 2018;**78**:5917–26.

Johnson A. Plotly: The front end for ML and data science models; 2012. https://plotly.com/ (4 April 2023, date last accessed).

Kasiviswanathan R, Copeland WC. Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase. *J Biol Chem* 2011;**286**:31490–500.

Klein HL. Genome instabilities arising from ribonucleotides in DNA. *DNA Repair (Amst)* 2017;**56**:26–32.

Koh KD, Balachander S, Hesselberth JR *et al.* Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nat Methods* 2015;**12**:251–7.

Koyanagi E, Kakimoto Y, Minamisawa T *et al.* Global landscape of replicative DNA polymerase usage in the human genome. *Nat Commun* 2022;**13**:7221.

Matsakis ND, Klock FS. The rust language. *Ada Lett* 2014;**34**:103–4.

Momjian B. *PostgreSQL: Introduction and Concepts*. Vol. **192**. New York: Addison-Wesley, 2001.

Moss CF, Dalla Rosa I, Hunt LE *et al.* Aberrant ribonucleotide incorporation and multiple deletions in mitochondrial DNA of the murine MPV17 disease model. *Nucleic Acids Res*2017;**45**:12808–15.

Nick McElhinny SA, Watts BE, Kumar D *et al.* Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A* 2010;**107**:4949–54.

Rawat P, Mahajan AN. ReactJS: a modern web development framework. *Int J Innov Sci Res Technol* 2020;**5**:698–702.

Reijns MAM, Parry DA, Williams TC, Colorectal Cancer Domain UK 100,000 Genomes Project *et al.* Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature* 2022; **602**:623–31.

Sassa A, Yasui M, Honma M *et al.* Current perspectives on mechanisms of ribonucleotide incorporation and processing in mammalian DNA. *Genes and Environ* 2019;**41**:7.

Sriramachandran AM, Petrosino G, Méndez-Lago M *et al.* Genome-wide nucleotide-resolution mapping of DNA replication patterns, single-strand breaks, and lesions by GLOE-seq. *Mol Cell* 2020;**78**: 975–85.e7.

The Launchbadge Team: The rust sql toolkit, 2023. https://github.com/launchbadge/sqlx (4 April 2023, date last accessed).

The Tokio Team: A runtime for writing reliable asynchronous applications with Rust, 2023. https://github.com/tokio-rs/tokio (4 April 2023, date last accessed).

Williams JS, Lujan SA, Kunkel TA *et al.* Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat Rev Mol Cell Biol* 2016;**17**:350–63.

Williams JS, Smith DJ, Marjavaara L *et al.* Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. *Mol Cell* 2013;**49**:1010–5.

Xu P, Yang T, Kundnani DL *et al.* Light-strand bias and enriched zones of embedded ribonucleotides are associated with DNA replication and transcription in the human-mitochondrial genome. *Nucleic Acids Res* 2024;**52**:1207–25.

Xu P, Storici F. Frequency and patterns of ribonucleotide incorporation around autonomously replicating sequences in yeast reveal the division of labor of replicative DNA polymerases. *Nucleic Acids Res* 2021a;**49**:10542–57.

Xu P, Storici F. RESCOT: restriction enzyme set and combination optimization tools for rNMP capture techniques. *Theor Comput Sci* 2021b;**894**:203–13.

Xu P, Storici F. RibosePreferenceAnalysis: analyzing the preference of rNMPs embedded in genomic DNA. *Software Impacts* 2021c; **10**:100149.

Zhou Z-X, Lujan SA, Burkholder AB *et al.* Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat Commun* 2019;**10**:3992.

Bioinformatics Advances, 2024, 00, 1–5
https://doi.org/10.1093/bioadv/vbae063
Application Note