

¿Cómo
funciona?

El ATAC-seq, la llave para abrir los secretos del genoma

ATAC-seq, the key to unlock the genome secrets

Resumen

El genoma de cada célula es enorme, por lo que se encuentra hiper compactado dentro del núcleo celular en forma de cromatina. Sin embargo, no toda la información codificada en nuestro ADN es necesaria en todo momento, por lo que sólo se requiere que algunos de los genes estén activos en determinada célula y en cierta condición. La cromatina tiene sitios completamente inaccesibles por el nivel de compactación, mientras que las regiones que se encuentran accesibles corresponden a los genes necesarios para la célula. El ATAC-seq es una técnica utilizada para conocer todas las regiones accesibles del genoma celular.

Palabras clave: Cromatina, compactación del ADN, transposasa, secuenciación masiva, bioinformática.

Summary

The genome of each cell is huge and highly compacted within the nucleus as chromatin. However, not all DNA-encoded information is required simultaneously; only a subset of genes is activated depending on the cell type and specific conditions. Chromatin contains inaccessible regions due to its high compaction level, while accessible areas correspond to functionally active genes. ATAC-seq is a powerful technique for accurately identifying accessible regions of the genome, offering critical insights into gene regulation and chromatin organization.

Keywords: Chromatin, transposase, massive-sequencing, DNA-compaction, bioinformatics.

Jaqueline Hersch-González
Víctor Julián Valdés*

Instituto de Fisiología Celular, UNAM, CDMX, México.

**Autor para la correspondencia:
julian.valdes@ifc.unam.mx*

¿Cómo cabe el genoma de 4,600 millones de pares de bases en un núcleo de 6 micrómetros?

Muchos estamos familiarizados con el hecho de que nuestras células tienen un núcleo donde se almacena nuestro ADN, es decir, nuestro genoma. Pero sigue siendo impresionante recordar que la longitud de nuestro genoma sería de poco más de dos metros si lo extendiéramos de manera lineal [1]. Uno se preguntará entonces, ¿cómo cabe todo nuestro genoma en un núcleo que mide tan solo 6 micras?, es decir, ¿un sitio tres millones de veces menor! ¡Es como acomodar 40 km de hilo dentro de una pelota de tenis! La respuesta es la compactación: el ADN está enrollado alrededor de unas proteínas denominadas histonas, formando una estructura llamada nucleosoma (Figura 1). Los nucleosomas van uno tras otro, como si fuera un “collar de perlas moleculares”. Después los nucleosomas hacen una estructura helicoidal, conocida como fibra de 30 nanómetros (por su grosor) [1]. Es como si fueras enrollando al collar de perlas alrededor de una ramita. En este punto el genoma está tan compacto que cabe dentro del núcleo, y a esa estructura hiper compactada de ADN y proteínas se le llama **cromatina**.

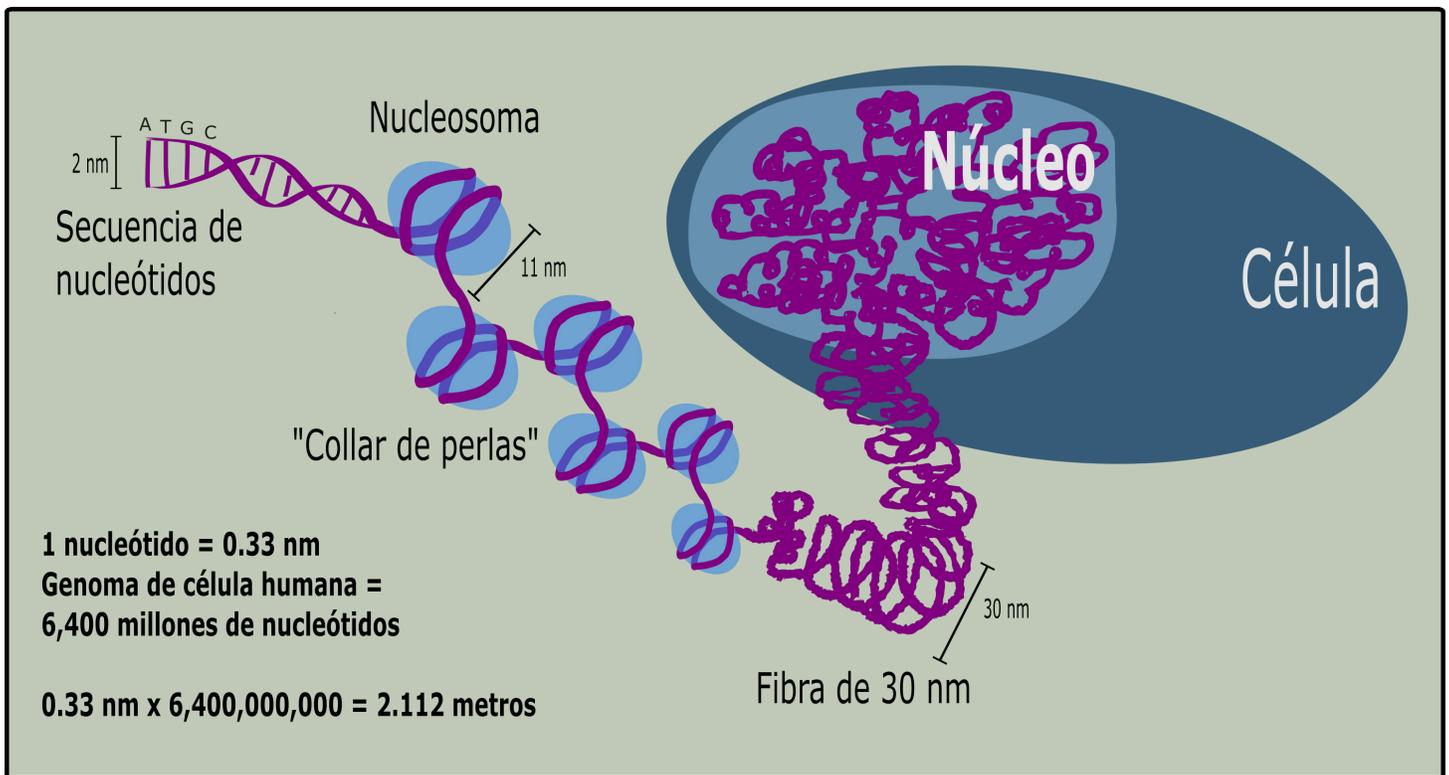


Figura 1. **Compactación de la cromatina dentro del núcleo eucariote.** La molécula de ADN se enrolla alrededor de los nucleosomas que están conformados por ocho proteínas histonas (H3, H4, H2A y H2B, dos de cada una) para formar la cromatina. Esta estructura se hipercompacta en la fibra de 30 nm dentro del núcleo.

¿Todas las células tienen la misma accesibilidad de la cromatina?

Aunque todas las células de nuestro cuerpo tienen la misma información genética, en él tenemos gran variedad de células y tejidos. Esto se debe a que las distintas células expresan distintos genes [2]. Por ejemplo, mientras que las neuronas expresan genes para sintetizar neurotransmisores, las células musculares expresan genes como la miosina, que nos permiten movernos. Conocer los mecanismos moleculares que permiten la “expresión diferencial” de genes entre tejidos, es una frontera de la biología moderna, y la compactación de la cromatina es uno de estos mecanismos.

Ahora que sabemos por qué es importante la compactación de la cromatina, nos interesa explorar una técnica que se usa para estudiar su accesibilidad. Pero, ¿a qué nos referimos con accesibilidad? Pensemos en un ejemplo: una secuencia de ADN que corresponde a un gen hipotético llamado “Genus”. Para que a partir del ADN de Genus se produzca una proteína, el gen necesita primero ser transcrito de ADN a ARN por la enzima “ARN polimerasa”, y

después ser traducido de ARN a proteína por medio de los ribosomas. Para que la ARN polimerasa tenga acceso al ADN de Genus, la secuencia necesita estar libre de nucleosomas o “accesible”. Regresando al ejemplo de los genes de neurotransmisores (neuronas) y miosina (músculo), la cromatina de estos genes debe estar accesible en las células donde se requieren, pero probablemente estén no accesibles en otros tipos celulares, donde estos genes no son necesarios. ¡Sorprendente como nuestras células optimizan los procesos moleculares!

Así pues, como ya puedes haber deducido, la cromatina se mantiene accesible en regiones donde hay genes importantes para que la célula se mantenga viva o lleve a cabo su función. A estos genes los podemos llamar “activos”, mientras que a los genes que no se expresan los podemos llamar “inactivos”. Por lo tanto, conocer los sitios de cromatina accesible del genoma de una célula nos puede dar un indicio de cuáles genes están activos.

Pero ahora, ¿cómo podemos saber cuáles zonas del genoma tienen cromatina accesible en una célula o tejido? Existen varias técnicas

que evalúan la accesibilidad, algunas tienen la desventaja de ser inespecíficas [3], pero una de las más eficientes y de mayor resolución es el “ensayo de cromatina accesible a transposasa utilizando secuenciación,” o ATAC-seq por sus siglas en inglés (*Assay for Transposase-Accessible Chromatin with sequencing*), que permite identificar las regiones de cromatina libres de nucleosomas [4].

¿Cómo funciona el ATAC-seq?

En el ATAC-seq utilizamos una proteína transposasa [4], esta proteína existe en casi todos los organismos y puede mover segmentos de ADN, conocidos como transposones, de un sitio a otro del genoma. Lo que más nos interesa de esta enzima es que tiene la capacidad

de cortar el ADN, y solamente puede hacerlo en regiones accesibles de la cromatina. En el ATAC-seq la transposasa se agrega a núcleos celulares intactos, pues queremos preservar la estructura de la cromatina tal cual está en un tejido o condición.

Para ayudar a nuestra imaginación veamos el esquema de la Figura 2, donde se representa la cromatina cerrada y la cromatina abierta (o accesible). La transposasa que se utiliza en ATAC-seq se llama *Tn5*, se produce en un laboratorio de manera artificial y tiene dos particularidades importantes: 1) es hiperactiva pues fue modificada por los científicos para aumentar su eficiencia; y 2) tiene adaptadores de ADN unidos a su estructura (verdes y rojos en la figura). Los adaptadores son secuencias cortas de

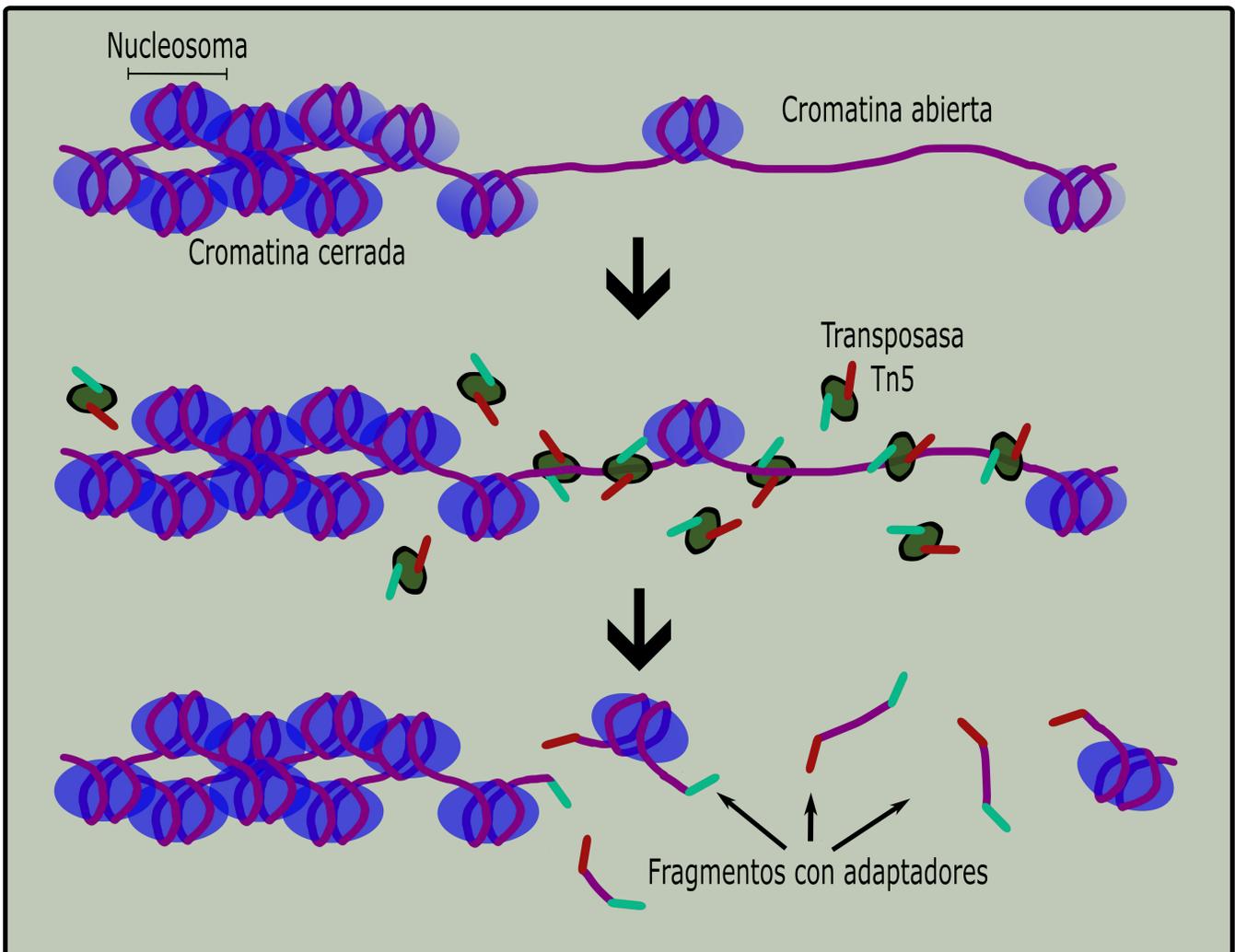


Figura 2. Esquema de la técnica de “ensayo de cromatina accesible a transposasa utilizando secuenciación” o ATAC-seq por sus siglas en inglés (*Assay for Transposase-Accessible Chromatin with sequencing*). Los núcleos celulares intactos son incubados con la transposasa hiperactiva *Tn5* que se produce de manera artificial. Esta transposasa corta en regiones accesibles e introduce adaptadores, que se representan en rojo y verde.

ADN sintético (~20 nucleótidos) y su función es importante en la amplificación y secuenciación. Una vez que entra en contacto con la cromatina, la Tn5 se une y corta el ADN en regiones accesibles e inserta también los adaptadores de ADN en cada uno de los extremos. Después de esto ya tenemos las regiones accesibles cortadas y marcadas con secuencias conocidas de ADN (adaptadores), ¡ya sólo queda identificarlas!

¿Y la secuenciación?

Después de incubar los núcleos con la Tn5, tenemos una mezcla de fragmentos de todo el genoma que corresponden a las regiones accesibles y que tienen unidos los adaptadores en sus extremos. Como quizás hayas escuchado antes, para saber qué nucleótidos conforman una molécula de ADN y su orden, se realiza un proceso llamado secuenciación. La secuenciación que se utiliza en el ATAC-seq es de las llamadas de “última generación” o NGS (*Next-Generation Sequencing*) y es tan eficaz que pueden secuenciarse millones de fragmentos de ADN al mismo tiempo, aunque no entraremos en los detalles técnicos en este artículo.

Pero antes de la secuenciación hay un paso muy importante: la amplificación de nuestros fragmentos por PCR, lo que se conoce también como preparación de bibliotecas. Cuando la Tn5 corta el ADN, inserta cada adaptador en lados opuestos, haciendo que los fragmentos queden desapareados en sus extremos, por lo que es necesario realizar la amplificación por PCR para

completar los extremos y que queden “parejos”. La otra parte importante de la amplificación es la de añadir los *barcodes* o índices, que son otra secuencia corta de ADN para diferenciar una muestra de otra (ej. tratamiento de alta glucosa vs baja glucosa). Por lo tanto, cuando se realiza la amplificación es crucial poner diferentes índices a distintas muestras (Figura 3).

Los adaptadores son importantes para la secuenciación porque son esenciales para la unión de nuestras bibliotecas a las celdas de flujo que se utilizan en la secuenciación masiva. Las celdas de flujo tienen adaptadores complementarios a los de nuestros fragmentos, lo que facilita su reconocimiento y unión a la celda.

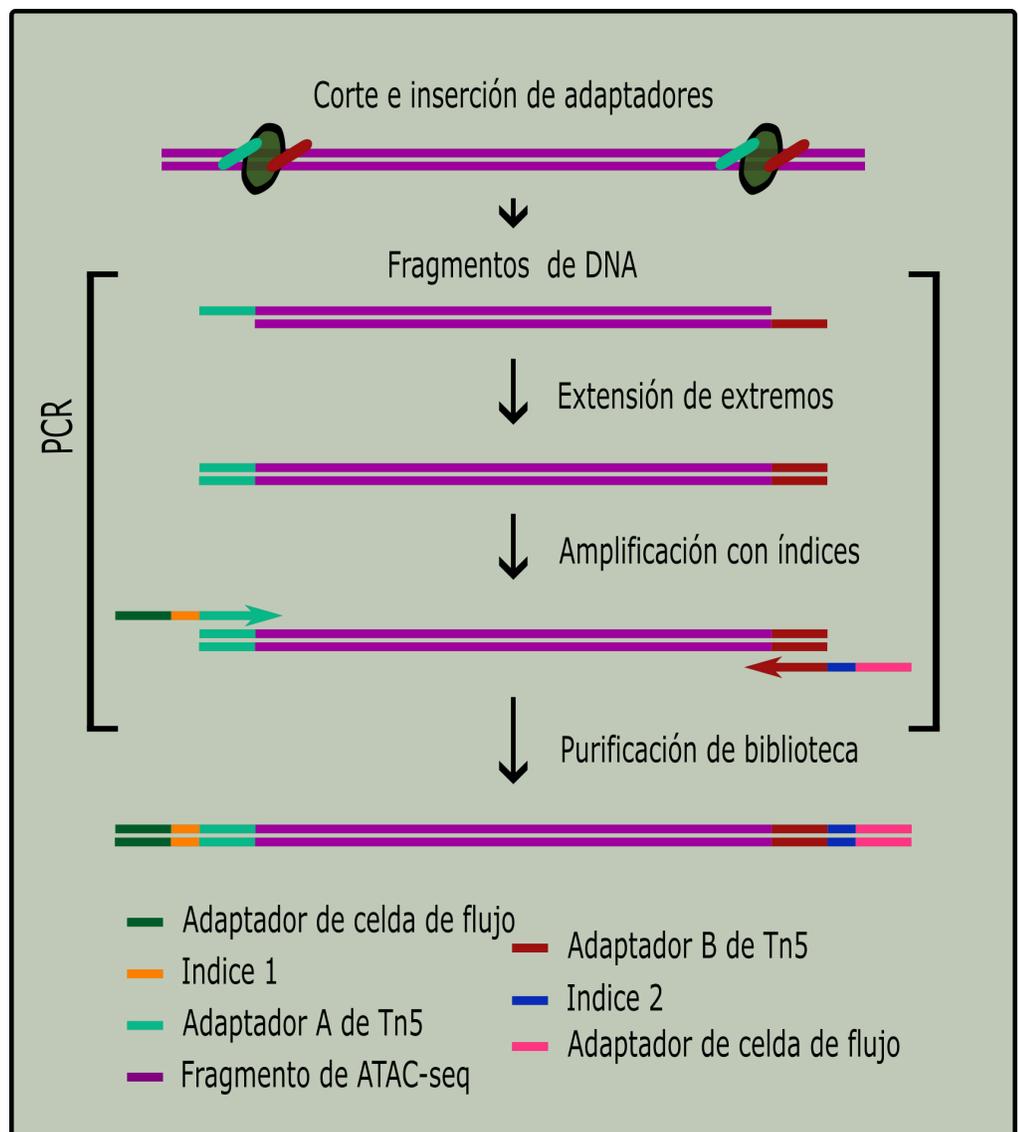


Figura 3. **Amplificación de bibliotecas de ATAC-seq.** Con el genoma fragmentado y con los adaptadores que puso la Tn5, se amplifican las librerías vía PCR completando los extremos y agregando las secuencias de adaptadores e índices para su identificación durante la secuenciación masiva de nueva generación.

Analizando todo el genoma: bioinformática

La secuenciación masiva genera miles de millones de secuencias de ADN que son imposibles de examinar a ojo, además en el ATAC-seq tenemos los fragmentos accesibles de TODO el genoma. Por ello se utilizan herramientas computacionales con capacidad de almacenamiento de varios terabytes para guardar, organizar y analizar estos datos.

Al analizar el ATAC-seq es importante eliminar las secuencias de los adaptadores que tanto nos sirvieron para que no interfieran con los siguientes pasos. Como trabajamos con todo el genoma, debemos alinear cada una de las secuencias, también llamadas lecturas, al genoma del organismo que estudiamos. Los organismos más investigados en el mundo tienen la secuencia de su genoma completo disponible en internet para cualquiera que lo necesite.

Y bien, después de varios pasos bioinfor-

máticos, podemos por fin generar gráficos y figuras para poder ver nuestros datos de una manera amigable. Si nos situamos en la Figura 4, podemos ver cómo las regiones libres de nucleosomas de “Genus”, que fueron fragmentadas por la Tn5, se alinean sobre el promotor, el cuerpo del gen y una región intergénica. Esa información es plasmada en la imagen como “picos de accesibilidad”.

Si comparamos dos muestras con condiciones distintas (ej. alta y baja temperatura), podríamos contrastar la dimensión de los picos en cada una y así saber las diferencias de la cromatina en cada muestra. Por ejemplo, en la imagen podemos ver que *Genus* está más accesible a temperaturas altas, por lo que biológicamente podríamos deducir que es importante para sobrevivir a esa temperatura y por eso está accesible. De hecho, en nuestro laboratorio estamos estudiando, mediante ATAC-seq, cambios en la accesibilidad de la cromatina de

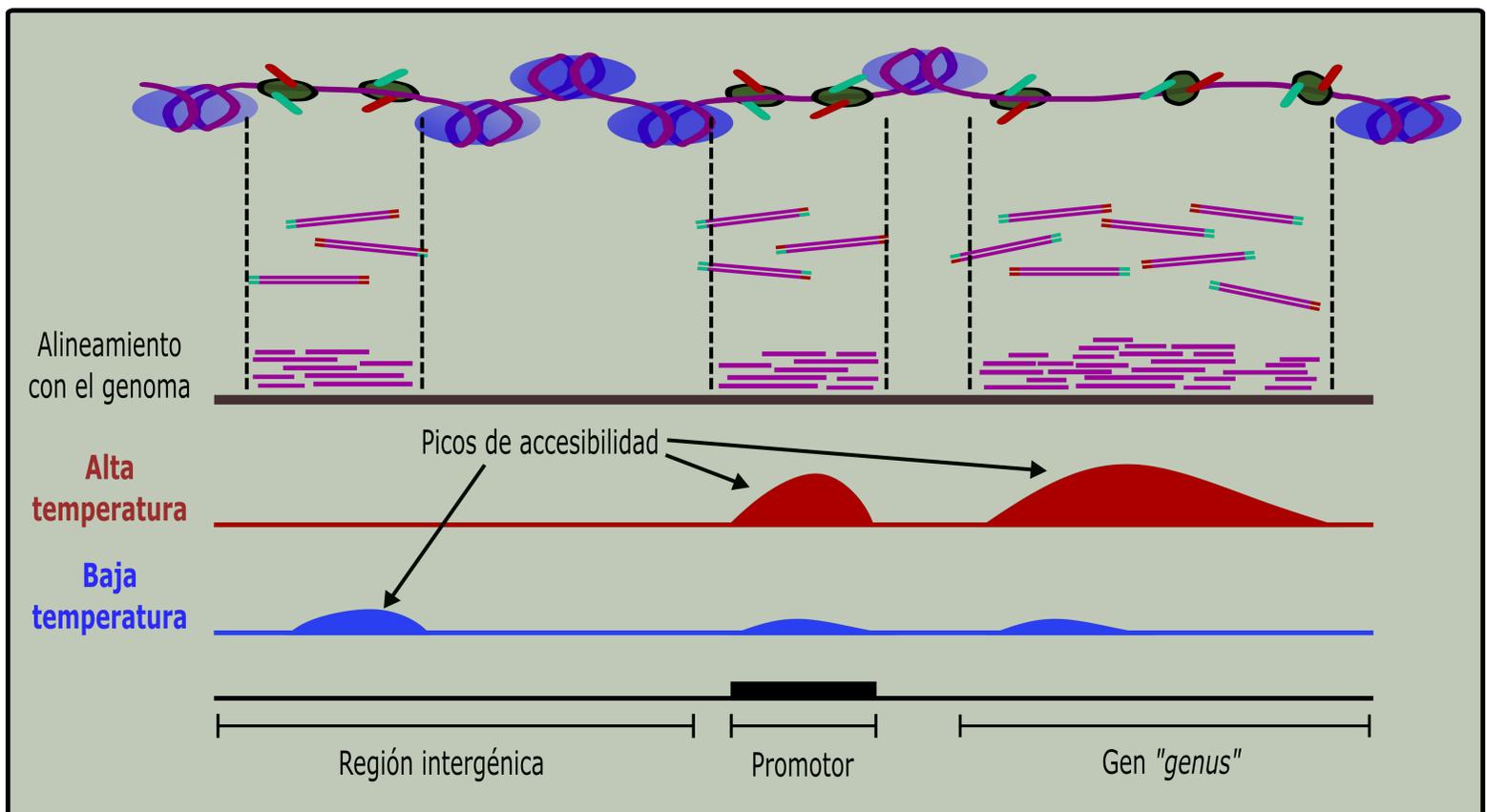


Figura 4. **Cambios en accesibilidad: alineamiento de lecturas al genoma y generación de “picos”.** Después de la secuenciación de las librerías de ATAC-seq, computacionalmente se alinean los millones de fragmentos secuenciados al genoma de referencia. Esto permite generar picos de accesibilidad y mapearlos a un gen de interés, por ejemplo el hipotético “Genus”, o a regiones intergénicas a lo largo del genoma. Al comprar dos (o más) ATAC-seq, podemos identificar cambios en la cromatina inducidos por distintos tratamientos, condiciones experimentales, entre células diferentes o, incluso, condiciones patológicas.

nemátodos *C. elegans* cuando son expuestos a altas temperaturas. También hemos identificado que algo tan simple como la glucosa, es capaz de inducir cambios permanentes en la accesibilidad de la cromatina de células humanas [5].

Más allá de la accesibilidad, ventajas y desventajas

Así pues, el ATAC-seq es una poderosa herramienta para analizar la accesibilidad de la cromatina de manera global, es decir, abarcando todo el genoma. Además, tiene la ventaja de requerir un muy bajo número de células para su implementación, unos cuantos miles es suficiente, e incluso algunas personas están realizando ATAC-seq a nivel de células individuales para diferenciar poblaciones de células específicas en una sola muestra [6]. Otra de las ventajas del ATAC-seq es que también evalúa regiones intergénicas, donde puede haber secuencias reguladoras distales (*enhancers*), que regulan la potencia de expresión de otros genes [6].

Ahora bien, no siempre mayor accesibilidad de la cromatina significa genes activos, a veces son necesarias otras señales para lograr la expresión de un gen. Por esto, el ATAC-seq se puede complementar con metodologías como el RNA-seq, que permite evaluar todos los ARN mensajeros de una célula, o el ChIP-seq, que evalúa cambios químicos en la cromatina.

No cabe duda que el ATAC-seq es una poderosa herramienta para conocer los cambios epigenéticos, es decir aquellos que alteran la expresión genética, sin alterar la secuencia del ADN. Con el desarrollo de este tipo de tecnologías ahora nos queda pensar, ¿qué preguntas nos gustaría responder con ATAC-seq? ¿Qué propuestas se te ocurren a ti? **iBIO**

Agradecimientos

Al programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM, Proyecto: IN217824 y CONAHCYT/SECIHTI (CB 0284867). Y al M. en C. Marlon Adrian Pulido Torres.

Referencias

- [1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2022). Chromosomal DNA and Its Packaging in the Chromatin Fiber. En: *Molecular Biology of the Cell*. 4th Edition. New York: Garland Science. Recuperado 27 de enero de 2025, de <https://www.ncbi.nlm.nih.gov/books/NBK26834/>
- [2] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2022). The Molecular Genetic Mechanisms That Create Specialized Cell Types. En: *Molecular Biology of the Cell*. 4th Edition. New York: Garland Science. Recuperado 27 de enero de 2025, de <https://www.ncbi.nlm.nih.gov/books/NBK26854/>
- [3] Parisi, N. (2023). Chromatin Structure Research Methods. *MATER METHODS* 2019; 9:2797. [dx.doi.org/10.13070/mm.en.9.2797](https://doi.org/10.13070/mm.en.9.2797)
- [4] Buenrostro, J., Wu, B., Chang, H., Greenleaf, W. (2016). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol.*; 109: 21.29.1–21.29.9. <https://doi.org/10.1002/0471142727.mb2129s109>
- [5] Wilson-Verdugo, M., Bustos-García, B., Adame-Guerrero, O., Hersch-González, J., Cano-Dominguez, N., Soto-Nava, M., Acosta, C.A., Tusie-Luna, T., Avila-Rios, S., Noriega, L.G., Valdes, V.J. (2024). Reversal of high-glucose-induced transcriptional and epigenetic memories through NRF2 pathway activation. *Life Science Alliance* 7(8), e202302382. <https://doi.org/10.26508/lsa.202302382>
- [6] Baek, S., Lee, E. (2020). Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. *Computational and Structural Biotechnology Journal*; (18):1429-1439. <https://doi.org/10.1016/j.csbj.2020.06.012>
- [7] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2022). How Genetic Switches Work. En: *Molecular Biology of the Cell*. 4th Edition. New York: Garland Science. Recuperado 27 de enero de 2025, de <https://www.ncbi.nlm.nih.gov/books/NBK26872/>
- [8] Yan, F., Powell, D.R., Curtis, D.J, Wong, N.C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biology*; 21:22. <https://doi.org/10.1186/s13059-020-1929-3>
- [9] Grandi, F.C., Modi, H., Kampman, L., Corces, M.R. (2022). Chromatin accessibility profiling by ATAC-seq. *Nature Protocols*. 1518-1552. <https://doi.org/10.1038/s41596-022-00692-9>