

1 | Basecalling

Basecalling es el proceso por el cual se convierte la señal eléctrica captada por los dispositivos de Oxford Nanopore en secuencias de nucleótidos. Esta conversión requiere de algoritmos computacionales avanzados, como las redes neuronales LSTM y los transformers utilizados por las opciones de basecalling actuales. Este proceso es crucial porque define la calidad de los datos, lo que afectará directamente los análisis posteriores.

Oxford Nanopore ha desarrollado múltiples basecallers a medida que ha habido avances tecnológicos. El basecaller actual es **Dorado**, el cual se puede utilizar mediante línea de comandos. Adicionalmente, está integrado en MinKNOW.

1.1 | Precisión del basecalling

Dorado cuenta con múltiples modelos de basecalling, cada uno con diferentes equilibrios entre precisión y requerimientos computacionales. Actualmente, los modelos disponibles son tres:

Modelo	Precisión Simplex	Requerimientos computacionales
fast [fast]	95.5%	
high accuracy [hac]	99.25%	
super accuracy [sup]	99.75%	

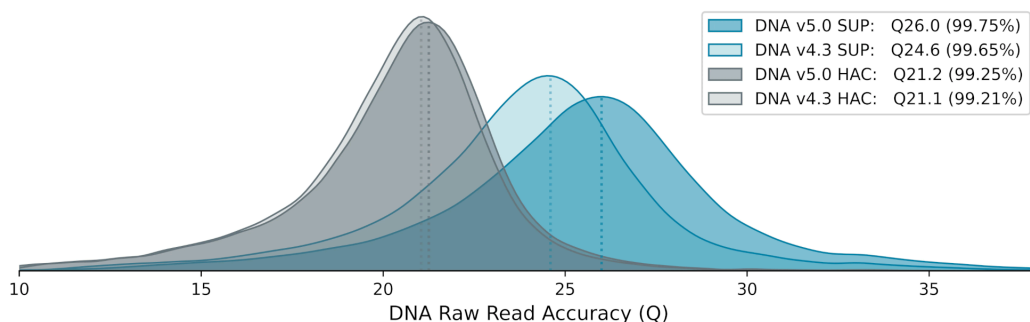


Figura 1. Precisión actual de los modelos de basecalling de Oxford Nanopore, usando el Kit V14 y celdas R10.4.1 de PromethION para secuenciación de genoma humano. Fuente: <https://nanoporetech.com/es/platform/accuracy>.

Usar el modelo sup es la opción recomendada si se busca obtener la máxima precisión posible. Sin embargo, este modelo tiene un costo computacional tan elevado que se vuelve impráctico de usar sin hardware dedicado (GPU).

1.2 | Uso básico de Dorado

`nanoporetech/dorado`

Dorado posee múltiples subcomandos con distintas funcionalidades, que incluyen basecalling, demultiplexación, descarga de modelos y otras.

1.2.1 | Basecalling

Para basecalling, se utiliza el subcomando `dorado basecaller`. La forma más básica de ejecutar este comando necesita el modelo a utilizar en el basecalling y el directorio con los archivos POD5. Por ejemplo, si tenemos el directorio `pod5/` y queremos utilizar el modelo `sup`, se ejecuta el siguiente comando:

```
dorado basecaller sup pod5/ > reads.ubam
```

sh

Adicionalmente, existen múltiples opciones adicionales que se pueden visualizar en la ayuda del comando.

1.2.2 | Demultiplexación

Para demultiplexar los archivos basecalled, se utiliza el subcomando `dorado demux`. Este comando requiere el archivo UBAM con los reads generado en el basecalling y el kit de barcoding. Por ejemplo, si tenemos el archivo `reads.ubam` y el kit SQK-NBD114-24:

```
dorado demux --output-dir basecalled_reads --kit-name SQK-NBD114-24 --emit-fastq reads.ubam
```

[sh](#)

En este ejemplo, `--output-dir` indica el directorio donde se guardarán los archivos demultiplexados, y `--emit-fastq` se utiliza para que los archivos generados estén en formato FASTQ (por defecto, se generan en BAM).