



# Data analysis for all levels of expertise

Accessible, scalable, workflows and tools

---

Jonghwa Ahn, Ph.D

Senior Field Applications Scientist

Oxford Nanopore Technologies





# Oxford Nanopore: from sample to answer

Comprehensive solutions for library preparation, sequencing and data analysis



## Prepare

- Field kits
- Lab kits
- Manual & Automated



## Sequence

- Field devices
- Lab devices
- Low & high output platforms  
(low \$ / test & low \$ / Gb)



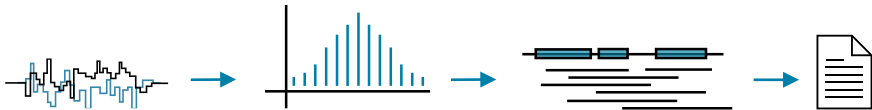
## Analyse

- Accessible
- Scalable
- Versatile



# Which sequencing data analysis is covered?

Oxford Nanopore provides a full range of solutions



|                 |                           |       | Bioinformatics expertise | OS                   | Basecall | QC | Analyse | Report |
|-----------------|---------------------------|-------|--------------------------|----------------------|----------|----|---------|--------|
| Full support    | MinKNOW                   |       | ● ● ●                    | Windows, Mac & Linux | ✓        | ✓  | ✓       | ✓      |
|                 | EPI2ME workflows*         | local | ● ● ●                    | Windows, Mac & Linux | ✓        | ✓  | ✓       | ✓      |
|                 |                           | cloud | ● ● ●                    | Windows, Mac & Linux |          | ✓  | ✓       | ✓      |
| Limited support | Research tools            |       | ● ● ●                    | Linux                | ✓        |    | ✓       |        |
|                 | Community developed tools |       | ● ● ●                    | Linux                | ✓        | ✓  | ✓       |        |

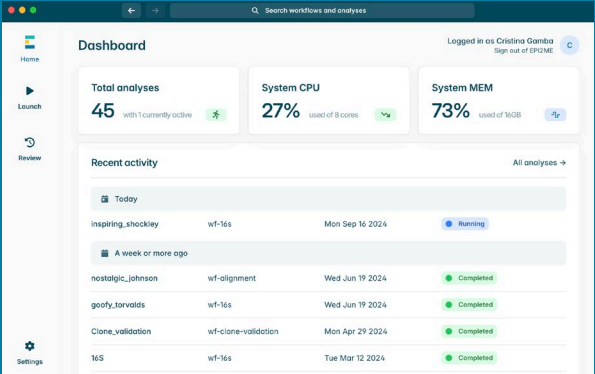
\* Accessible from both an intuitive interface and the command line



More at: <https://nanoporetech.com/data-analysis>

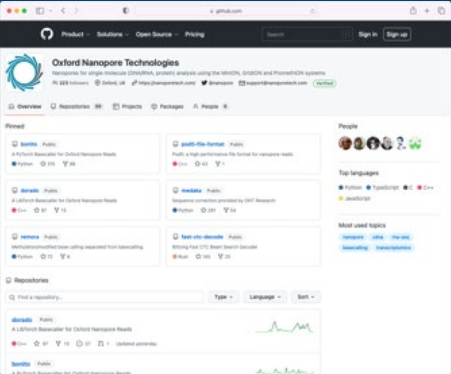


# Oxford nanopore data analysis options



The screenshot shows the EPI2ME dashboard. At the top, it says 'Dashboard' and 'Logged in as Cristina Gamba'. Key metrics include 'Total analyses: 45', 'System CPU: 27%', and 'System MEM: 73%'. Below these are sections for 'Recent activity' and 'All analyses', listing various workflow runs with their status (Running or Completed).

EPI2ME



The screenshot shows the GitHub repository page for Oxford Nanopore Technologies. It displays the repository name, description, and a list of recent commits with their authors and dates.

GitHub

|                   |  |  |
|-------------------|--|--|
| Location          | Local, distributed or in the cloud           | User defined (laptop, cluster or cloud)      |
| Timing            | Post-run or real time                        | Post-run                                     |
| Configurability   | Pre-configured                               | User defined                                 |
| Reporting         | Detailed output, shareable reports           | User built                                   |
| Operating systems | Windows, Mac, Linux                          | User defined                                 |
| Expertise needed  | <div><div></div><div></div><div></div></div> | <div><div></div><div></div><div></div></div> |



Download the EPI2ME app for free



Get the latest tools from GitHub





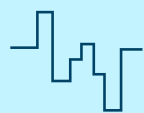
# Basecalling

- MinKNOW
- Dorado



# MinKNOW: overview

## Our devices software: from basecalling to reporting



### Basecalling

#### Basecalling

- Real-time basecalling
  - Choose desired accuracy
- Methylation calling



### QC

#### Real time QC

- Real-time quality control of the run
  - Stop and resume as needed



### Analyse

#### Analyse your data

- Seamless Analysis in real time including
  - Enrichment or depletion with Adaptive sampling
  - Alignment
  - Barcode demultiplexing
  - And more



### Report

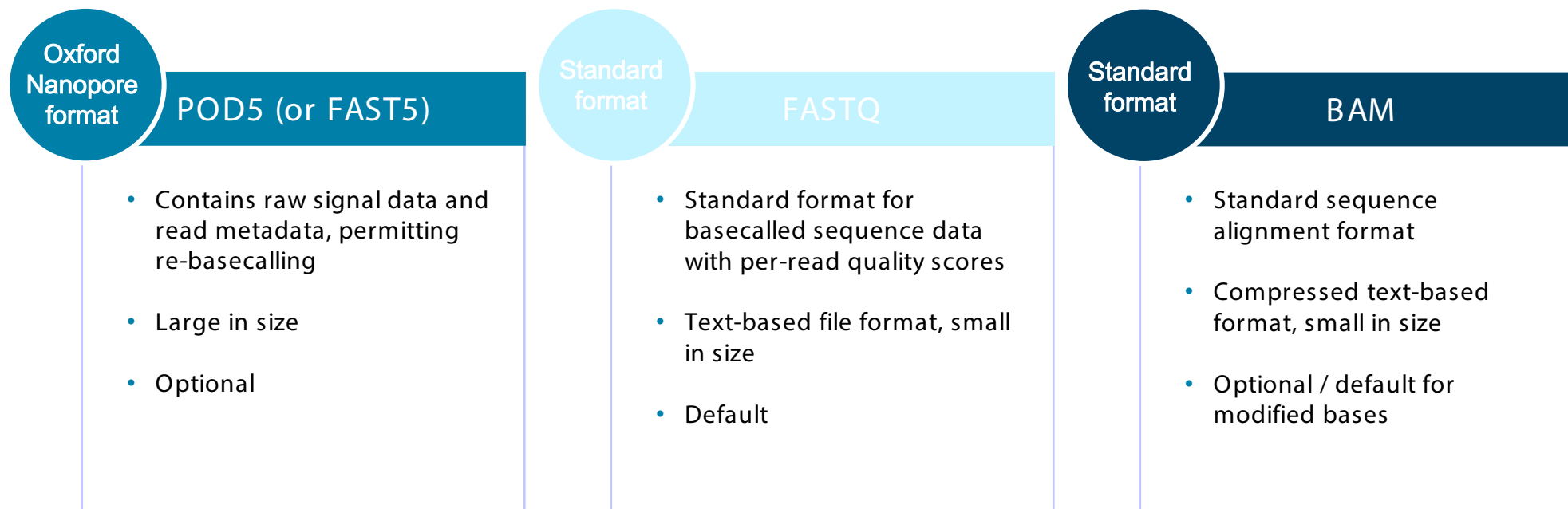
#### Generate automatic report

- QC reports of each run



# Store only the data you need

## Data formats





# Example storage requirements

GPU: V100 GPU  
Memory:64 GB RAM  
Storage: **4 TB SSD**



## Per flow cell storage examples



| Flow cell output | POD5<br>(optional<br>storage) | FASTQ.GZ    | Unaligned BAM<br>with<br>modifications |
|------------------|-------------------------------|-------------|--|
| 30 Gb            | 150 GBytes                    | 19.5 GBytes | 18 GBytes                              |
| 150 Gb           | 1 TByte                       | 97.5 GBytes | 90 GBytes                              |

## Full run storage examples

Assuming 30/150 Gbases per flow cell

| No. flow cells /<br>run    | POD5<br>(optional<br>storage) | FASTQ.GZ    | Unaligned BAM<br>with<br>modifications |
|----------------------------|-------------------------------|-------------|--|
| 5 MinION flow<br>cells     | 750 GBytes                    | 97.5 GBytes | 90 GBytes                              |
| 2 PromethION<br>flow cells | 2 TBytes                      | 195 GBytes  | 180 Gbytes                             |



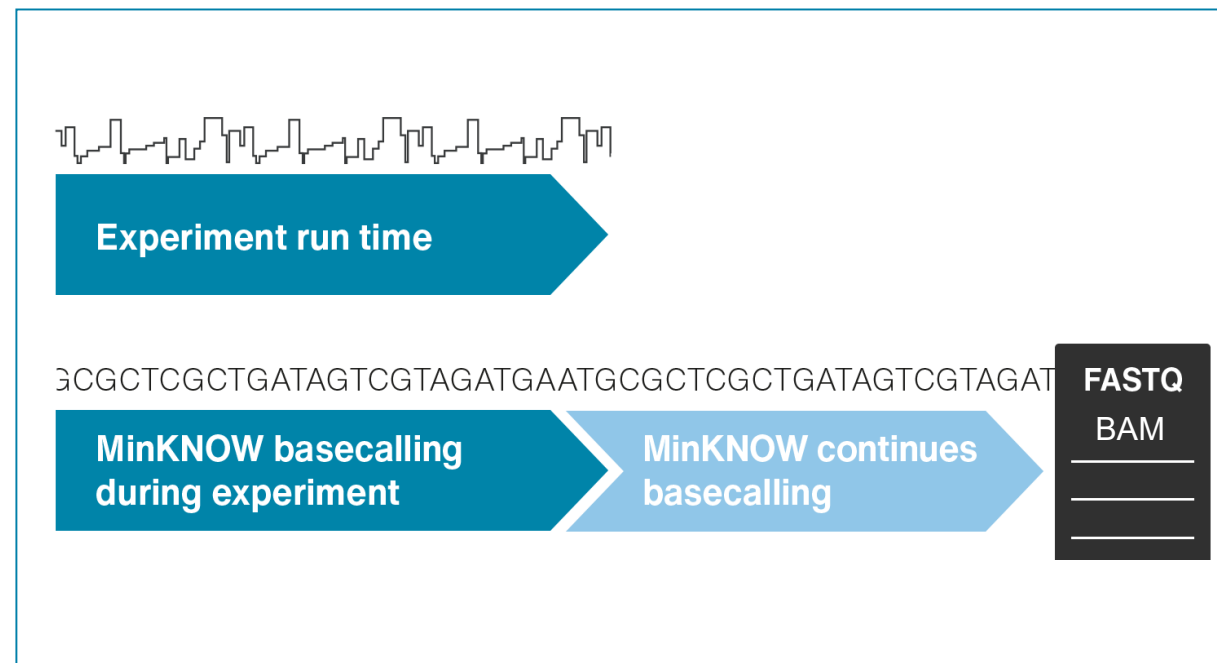
# Canonical basecalling

## From raw data to sequence

- Transform the “squiggle” into sequence with AI
- Basecall in real-time with MinKNOW
- Post-run basecalling with MinKNOW or Dorado
- Choose the most suitable model
  - HAC is faster, recommended for most applications
  - SUP is slower, recommended for *de novo* assembly projects or low-frequency variant analysis
- Methylation calling can be switched on with a click



Find out more at: [nanoporetech.com/how-it-works/basecalling](https://nanoporetech.com/how-it-works/basecalling)





# Inbuilt modification calling

## Most accurate methylation

- Seamless base modification analysis of native DNA data in MinKNOW
- In parallel to basecalling, real-time options
- Currently supports multiple DNA and RNA modification calling
- Latest models available in Dorado standalone



Find out more at: <https://nanoporetech.com/platform/accuracy>

### Basecalling

Basecalling model

High-accuracy basecalling ▾

Modified bases ?

☒ 6mA in all contexts

☒ C modifications

- ☒ 4mC and 5mC in all contexts
- ☐ 5hmC and 5mC in CG contexts
- ☐ 5hmC and 5mC in all contexts

Basecalling options in MinKNOW



# Canonical basecalling models and versions available

## In MinKNOW and Dorado

### Basecalling Models

|     | Model Version | Key Features                                      | Dorado | MinKNOW              |
|-----|---------------|---|--------|----------------------|
| DNA | v4.3          | Live basecalling with HAC                         |        |                      |
|     |               | Improved bacterial/low-complexity regions         | ✓      | ✓                    |
|     |               | Mods: 5mC/5hmC, 5mC, 6mA                          |        |                      |
|     | v5.0*         | Raw read accuracy ~Q26 SUP bacterial improvements | ✓      | ✓                    |
|     |               | Mods: 4mC/5mC                                     |        |                      |
|     | v5.2*         | Raw read accuracy ~Q21 HAC                        |        |                      |
|     | v5.2*         | Raw read accuracy ~Q22 HAC                        | ✓      | in following version |
| RNA | v3.0.1        | Raw read accuracy ~Q13 SUP<br>Mods: m6A DRACH     | ✓      | ✓                    |
|     | v5.0.0*       | Raw read accuracy ~Q19 SUP<br>Mods: m6A, pseU     | ✓      | ✓                    |
|     | v5.1*         | Mods: inosine, m5C                                | ✓      | ✓                    |
|     | v5.2*         | Mods: 2'Ome (A,C, U, G)                           | ✓      | in following version |

\* SUP transformer architecture

### Modified Bases

|     | Model                   | Context | Dorado | MinKNOW              |
|-----|-------------------------|---------|--------|----------------------|
| DNA | 5mC/5hmC (mammal/plant) | CpG     | ✓      | ✓                    |
|     |                         | All     | ✓      | ✓                    |
|     | 4mC/5mC (bacterial)     | All     | ✓      | ✓                    |
|     |                         | All     | ✓      | ✓                    |
| RNA | m6A                     | DRACH   | ✓      | ✓                    |
|     |                         | All     | ✓      | ✓                    |
|     | PseU                    | All     | ✓      | ✓                    |
|     | m5C                     | All     | ✓      | ✓                    |
|     | Inosine                 | All     | ✓      | ✓                    |
|     | 2'Ome-A                 | All     | ✓      | in following version |
|     | 2'Ome-C                 | All     | ✓      | in following version |
|     | 2'Ome-U                 | All     | ✓      | in following version |
|     | 2'Ome-G                 | All     | ✓      | in following version |
|     |                         |         |        |                      |

\* Currently only available in SUP



# Accuracies for v5.0 modified base models

In MinKNOW 25.05.14

| Modified base model | Context | HAC accuracy | SUP accuracy |
|---------------------|---------|--------------|--------------|
| 4mC and 5mC         | All     | 97.1%        | 98.0%        |
| 5mC and 5hmC        | CpG     | 99.3%        | 99.3%        |
|                     | All     | 98.6%        | 98.7%        |
| 6mA                 | All     | 97.5%        | 98.3%        |

Benchmarking Modified Base Detection with Modkit in Oct. 2024

| Modified Base | Context | HAC Accuracy | SUP Accuracy |
|---------------|---------|--------------|--------------|
| 5mC+5hmC      | All     | 97.30        | 97.80        |
| 5mC+5hmC      | CpG     | 98.21        | 98.15        |
| 5mC only      | All     | 99.20        | 99.48        |
| 5mC only      | CpG     | 99.76        | 99.81        |
| 6mA           | All     | 96.21        | 97.60        |





# Blackwell Architecture GPU Support

In MinKNOW 25.05.14

|   |                   | HAC basecalling<br>only | HAC basecalling<br>with alignment | SUP basecalling<br>only | SUP basecalling<br>with alignment |
|---|-------------------|-------------------------|-----------------------------------|-------------------------|-----------------------------------|
| Standalone<br>MinKNOW on<br>NVIDIA GeForce<br>RTX 5090 - MinION<br>Mk1D       | DNA (30 kb human) | 1                       | 1                                 | 1                       | 1                                 |
|   | RNA               | 1                       | 1                                 | 1                       | 1                                 |
| Standalone<br>MinKNOW on<br>NVIDIA GeForce<br>RTX 5090 -<br>PromethION 2 Solo | DNA (30 kb human) | 2                       | 2                                 | 0.8                     | 0.8                               |
|   | RNA               | 2                       | 2                                 | 0.8                     | 0.8                               |

- 30 Gb output over 72 hours from a MinION flow cell running 30 kb human reads
- 100 Gb output over 72 hours from a PromethION flow cell running 30 kb human reads
- 7 Gb output over 72 hours from a MinION flow cell running RNA
- 35 Gb output over 72 hours from a PromethION flow cell running RNA



# Sequence data types



# File formats: FASTA

The reference file

## Notes

- Header always begins with >
- Unique Identifier followed by optional information separated by a space
- Followed by lines of sequence
- Required input for many applications

## Often used for read alignments:

- Reference bias
- Population representation
- Species diversity
- Strain/breed/sub-species relatedness
- Not representative of any one member

Unique identifier

Optional metadata

```
>XI dna:chromosome chromosome:R64-1-1:XI:1:666816:1 REF
CACCACACCCACACACCACCCACACACACACCACCCACACACCACCCACACACC
ACACCCACTACTCTAACCCTATTCTAATCCAACCCTGATCAACCTGTCTCCAAACCTACC
CTCACATTACCCTACCTCTCCACTCGTTACCCTGCCCCACTCAACCATAACCTCCCATT
CACCATCCATCTCTCTACTGTCACCAGCCCACCGTCCACCATAACCGTTACCCTCCCATT
ACTCATATTTAACCCCACTACCACTTACCCTGCCATTACCCTACCATCCACCATGTCCTA
CTCACTGTACTGTTGTTCTACCCTCCATATTGAAACGTTAACAATAATCGTAAATAATA
CACATATACTTACCCTACCACTCTAATCCCACCACACATCACATGCCATACTCACCTTCA
CTTGATACTGATATGGTATACGCACACGGATGCTACGTATATACCACTCTCAACTTACC
CTACTCTCACATTCCACTCCATGGCCCAGTCTCACTAAATCAGTACGATGCACTCACATC
ATTATTCACGGCACTTGCCTCAGCGGTTTATACCCTGTGCAATTTACCCATAAAACCCAC
GATTATCCACATTTTAATATCTATATCTCATTACGCGGCTCCAAATATTGTATAACTGCT
CTTAATGCATACATTATACCACTTTTACTCCATATACTAACCATTCAATTTATACACACT
TATTTCAATATACCCACAAAATCACCCTAAAATTACCTAAACATAAAAAATATTCTACTC
TTCAACAATAATACATAAACACACTCAATTGCGTATCTATACCACCATGACGTCATTAAC
```

Reference sequences provide coordinate systems to anchor our prediction against

This facilitates the comparison of predictions derived from different samples and studies



# File formats: FASTQ

- De facto format for most sequence data types
- Standard text file (huge in size)
- Blocks of 4 lines correspond to data for 1 read
  - Header, sequence, header (denoted by "+"), per-base quality

Diagram illustrating the FASTQ file format structure with annotations:

- @Label**: Points to the first line of the header block.
- Read ID**: Points to the second line of the header block.
- Run ID**: Points to the third line of the header block.
- Sample ID**: Points to the fourth line of the header block.
- Sequence**: Points to the sequence line (the line starting with 'G').
- Q scores represented as ASCII characters**: Points to the quality score line (the line starting with '+').

Example FASTQ record:

```
@64107149-aefe-4445-98dd-63b746dd6d08 runid=12b8e1a5cfeab8afe4cc42710993b762f9fdae5 sampleid=DM_lambda read=845 ch=149 start_time=2018-10-09T18:39:45Z
CGGTATTGCTTCGTTCAAGTATTGCTTCAACGACGTTCTCGGTTTCATCGGAGGATGGAGTGAAAGAGATGCGCTATTACGAAAAAATTGATGGCAGCAAATACCGAATATTTGGGTGGTTGGCGATCTGCACGGATGCTACACGAACC
+
, & % & & % # ' * 7CF7C$@=>@@F5C9?85IDGHA;?>99E7?+=* .&18>9:(&+('5;;<?A6879'D347*/?=:CFJE=@;%,0BEG@'AJB;?98-, ,<<2-*)& .-23344259%$&.0310//:@>A=9=AB?<:7651D6:E:FD
@cda0e6e0-5e0c-4042-911f-d917d2ef94c8 runid=12b8e1a5cfeab8afe4cc42710993b762f9fdae5 sampleid=DM_lambda read=927 ch=8 start_time=2018-10-09T18:36:01Z
CCGTAGCCTTCGTTCAAGTATTGCTGGCGGTATATTTCTCCAGCGGCGCTCTGCGGCCGTTTCGTAAGCCTTCTGCGCCTCTTCGGTATATTTAGCCGTGACCTTCGGTATCGGCGCTCTGCTGCTGCGCGTCTTTTGTCTCTGTTAG
+
+$$$#%$#)-)065<-842.$*, -)++ ,0;5/5;CDAA@IEBD9D=*78*-, .74&$*$1)' : ;=1589' -+---468GCDDC=B637@D DDFE;/CB9<C?%('(<&+>:7=,+'479=:6:.'),(%%&##*+,--589DBD/8&2
@d2352a8b-2331-413c-902f-8ef920fa82cd runid=12b8e1a5cfeab8afe4cc42710993b762f9fdae5 sampleid=DM_lambda read=954 ch=310 start_time=2018-10-09T18:35:36Z
GTTATTACTTCGTTCAAGTATTGTAGGTCGCCCCTAACCTGTCAAGTACCAGGAAGGGACCAGTAAAGAGAGATAATGATTATGTCTACATATCTGGCGTAACGTGCGTGGAGCCATCAAACCGTCAAATAATCAATTATGACGCAGGT
+
)22&&((%-%/(<:@8=?A0232303357...*$'&.14A2AC=727//, %(&*+% '$, %(+%%&(1AA2%%6:AB5PECGEB@?::;D;5?DB/*? '&#$#"(/?>>>- .&&&%*<=-*:98843%0;I@AEGT=6: , , *2<<DB>CG6=@>
```

Legend:

- G** → Base G
- +** → Q="+"= 10



# File Formats: SAM & BAM

## SAM files

- Sequence alignment map (SAM) first published by Li *et al.*, 2009
- File specification: <https://www.htslib.org/doc/sam.html>
- SAM contains one line for each read and has 12+ columns containing:
  - Coordinates, mapping quality, and metrics describing each read alignment
- Reads can also be stored unaligned to a reference genome

## BAM files

- BAM file is the binary (computer readable) version of a SAM
- Same information
- Much smaller storage space required
- Can index BAMs for rapid, random access
- If you have SAM, you should convert to BAM

Read ID      Alignment info      Query sequence

```

@HD VN:1.6 SO:unknown
@PG ID:basecaller PN:dorado VN:0.5.2+7969fab3 CL:dorado basecaller hac pod5/DNA
@PG ID:samtools PN:samtools PP:basecaller VN:1.19 CL:samtools view -h calls.bam
@RG ID:94b5ed537ad8f50ca680ab01ecb5efe3c60a1f81 dna r10.4.1 e8.2 400bps hacv4.3.0 PU:RAX12777 PM:Stevens-MacBook-Pro.lo
a8a2d9b4-2d86-4899-8b37-fef8e5d4c9c5 4 * 0 0 * * 0 0 AAGGTTAAACGTAACCTGGTTTGTTCCTGAACAGCACCTAAGTTTGAT
db99d697-ee48-47ea-80cf-f5c7fddc9a1f 4 * 0 0 * * 0 0 TCCTACTCGTTCATTACGTATTGCTAAGGTTAAGTGTAGTTGGTTTG
3541d563-b3b3-4423-b5e7-709b478a75ca 4 * 0 0 * * 0 0 CCATTAACCTGGTTCGTTACGTATTGCTAAGGTTAATAGTTGGATGACC
99a35303-86ef-414a-9b2f-2d5123b31a5b 4 * 0 0 * * 0 0 AAGGTTAACCAGTAGAAGTCCGACACATCAGATCAGCACCTTCCCGG
a95fd88c-0a8c-4151-bf6c-89d771a781ab 4 * 0 0 * * 0 0 TGTATGCTGGTAACCTTGCTTCATTACGCGTATTGCTAAGGTTAAGCA
02e53d68-6c54-4945-9174-e83b311acb06 4 * 0 0 * * 0 0 GTTATGTATACACTAATTGGTCTGTGTTAGGTTCCATAACTCGCTGGC
7323297e-2ee3-4ad5-ab66-3b912ae371cf 4 * 0 0 * * 0 0 AAGGTTAACGTCACCTGACAGTGGTTCGTAAGGTTAAGGTTGATGATGA
4526f733-23d8-4c77-935e-117b3eb4db0e 4 * 0 0 * * 0 0 ATAACCTACTTGGTACCAGTACGTATTGCTAAGGTTAATAGTTGGATGA
661ace18-5e5d-4a2b-9b65-74c17ee4f7a8 4 * 0 0 * * 0 0 AAGGTTAAGCATAGTTCTCAGTGATGGGGGCACAGCACCTAGAGTTTGA
aad8e1a7-03fa-429c-89d7-8dc7417a478f 4 * 0 0 * * 0 0 GTGGCCCGCTTGGTTCAGTTACGTATTGCTAAGGTTAAGGTTAAGGTTGAC
ff5ea7e1-e7a3-4bb3-b80b-60ccffc026c5 4 * 0 0 * * 0 0 AAGGTTAACCAGTAGAATCCGACAACTCATCAGCACCTAGAGTTTGA
6dfa22c1-1f25-47d5-96f8-ad2d2d99075b 4 * 0 0 * * 0 0 ATGGTCAGCAGATTACGTATTGCTAAGGTTAAGGTTAAGGTTTGGTTGTT
f58369de-6d1c-490d-a5d5-6471b7cd49c6 4 * 0 0 * * 0 0 AAGGTTAATAGTTGGATGACCAAGGATAGCCAGCACCTTACGGTTAC
8a6f9ea2-4229-4cde-beb7-203d3ef37e46 4 * 0 0 * * 0 0 GTGTAACTACTGGTTCAGTTAGTATGCTAAGGTTAAGGTTAAGGTTGTT
c3186027-c8cd-4aca-8cab-9473ae9bdfad 4 * 0 0 * * 0 0 TTGTTTTGCTGGTTCAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
b252f741-f368-4353-aa71-cd5a0d5ea3f6 4 * 0 0 * * 0 0 ATGTTATGAGCCTTAAGCTTCAGTTCTGCATATACATATTCATATGAGC
32eace3b-de7d-471c-9d7f-10aa6ace13a7 4 * 0 0 * * 0 0 ATATCCGTATTGCTTCGTTTCAGTTACATATACATACGAGCGACAGCCCA
4f9dd10e-2923-4955-81af-0369167b83ae 4 * 0 0 * * 0 0 AAGGTTAAACGTAACCTGGTTTGTTCCTGAACAGCACCTAGAGTTTGA
f4844cb6-cb9d-4db8-81f5-d8afc2750db2 4 * 0 0 * * 0 0 ATGTTGCTGTACTTGGTTCGTTGCTGTTGCTAAGGTTAAGGTTAGTACTG
034fd862-e83e-48e7-a136-c3d5115a00e0 4 * 0 0 * * 0 0 TTGTGTATTCTTCATCTATTATAAGGTTAAGGTTAAGGTTAAGGTTAAGGTTA
d393ceaf-f691-4cad-b2f9-5047fb03a38f 4 * 0 0 * * 0 0 ATGTACATATTTATTTCGTTTCAGTTTCGTTCTGTGCTAAGGTTAGACGTT
176ed11f-09a4-45e7-8e0e-7c7e7bde250f 4 * 0 0 * * 0 0 AAGGTTAATAGTTTGGATGACCAAGGATAGCCGACGACCTAGAGTTTGA
  
```



# Methylation detection in MinKNOW

Ensure BAM output is enabled to store methylation information

BAM can be used to store aligned or unaligned reads (referred to as a uBAM)

- BAM or uBAM required by some EPI2ME Labs workflows (e.g. wf-human-variation)
- Methylation information described in the MM and ML tag of the BAM file

**Output**

Data saved as  
Test/

Output location  
/data/

Output format  
☒ FAST5 ☒ FASTQ ☒ **BAM**

☒ Filtering

Qscore: 9 | Readlength: Unfiltered | Read splitting: Disabled **Options**

▼ Advanced options

**FAST5 is a raw data format**

- Signal data

**FASTQ is a common format for storing sequencing data**

- Your actual reads and associated per-base quality information

**No reference file specified**

- Unaligned BAM file with MM and ML tags

**Reference file specified**

- Aligned BAM file with MM and ML tags and reads mapped to reference genome





# Generating an alignment

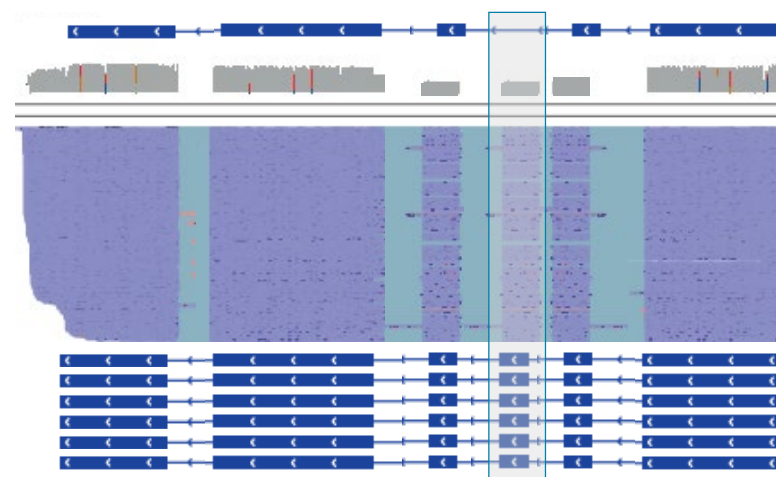
Required first step for most analyses

- For a sequencing read, to determine the location of origin within the reference sequence
- Reference can be complete genome, transcriptome or de novo assembly
- Input is typically FASTQ and reference genome in FASTA format
- Output is sequence alignment map (SAM) format file or binary equivalent (BAM)



## Genomics

SNPs/indels  
Structural variation  
Repeat content  
...and more



## Transcriptomics

Expression  
Isoforms  
Variation  
...and more



# Where to find reference sequences?

## Public databases

### Ensembl

- Database of sequences and corresponding annotations
- [ensemblgenomes.org](https://ensemblgenomes.org)

### UCSC

- Database of sequences and corresponding annotations
- Genome Browser
- [ucscbrowser.genenetwork.org](https://ucscbrowser.genenetwork.org)

### RefSeq

- Database of sequences
- [ncbi.nlm.nih.gov/refseq/](https://ncbi.nlm.nih.gov/refseq/)

| Name   | Last modified    | Size | Description |
|--|------------------|------|-------------|
| Parent Directory                             | -                | -    | -           |
| <a href="#">acanthochromis_polyacanthus/</a> | 2018-09-06 04:10 | -    | -           |
| <a href="#">ailuropoda_melanoleuca/</a>      | 2018-09-06 05:21 | -    | -           |
| <a href="#">amphilophus_citrinellus/</a>     | 2018-09-06 10:52 | -    | -           |
| <a href="#">amphiprion_ocellaris/</a>        | 2018-09-06 09:49 | -    | -           |
| <a href="#">amphiprion_percula/</a>          | 2018-09-06 11:31 | -    | -           |
| <a href="#">anabas_testudineus/</a>          | 2018-09-06 13:13 | -    | -           |
| <a href="#">anas_platyrynchos/</a>           | 2018-09-05 08:26 | -    | -           |
| <a href="#">anolis_carolinensis/</a>         | -                | -    | -           |
| <a href="#">aotus_nancymae/</a>              | -                | -    | -           |
| <a href="#">astatotilapia_calliptera/</a>    | -                | -    | -           |
| <a href="#">astyanax_mexicanus/</a>          | -                | -    | -           |
| <a href="#">bos_taurus/</a>                  | -                | -    | -           |

| Name   | Last modified    | Size | Description |
|--|------------------|------|-------------|
| Parent Directory   | -                | -    | -           |
| <a href="#">CHECKSUMS</a>  | 2018-09-12 16:22 | 3.8K | -           |
| <a href="#">README</a>   | 2018-09-03 22:12 | 4.9K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.I.fa.gz</a>    | 2018-09-03 22:12 | 70K  | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.II.fa.gz</a>   | 2018-09-03 22:12 | 248K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.III.fa.gz</a>  | 2018-09-03 22:12 | 97K  | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.IV.fa.gz</a>   | 2018-09-03 22:11 | 465K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.IX.fa.gz</a>   | 2018-09-03 22:12 | 134K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.Mito.fa.gz</a> | 2018-09-03 22:12 | 22K  | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.V.fa.gz</a>    | 2018-09-03 22:12 | 176K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.VI.fa.gz</a>   | 2018-09-03 22:12 | 83K  | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.VII.fa.gz</a>  | 2018-09-03 22:11 | 333K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.VIII.fa.gz</a> | 2018-09-03 22:12 | 171K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.X.fa.gz</a>    | 2018-09-03 22:12 | 227K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.XI.fa.gz</a>   | 2018-09-03 22:12 | 204K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.XII.fa.gz</a>  | 2018-09-03 22:12 | 324K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.XIII.fa.gz</a> | 2018-09-03 22:12 | 282K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.XIV.fa.gz</a>  | 2018-09-03 22:12 | 239K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.XV.fa.gz</a>   | 2018-09-03 22:11 | 333K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.XVI.fa.gz</a>  | 2018-09-03 22:12 | 289K | -           |
| <a href="#">Saccharomyces_cerevisiae.R64-1-1.dna.toplevel.fa.gz</a>        | 2018-09-03 22:12 | 3.6M | -           |





# EPI2ME

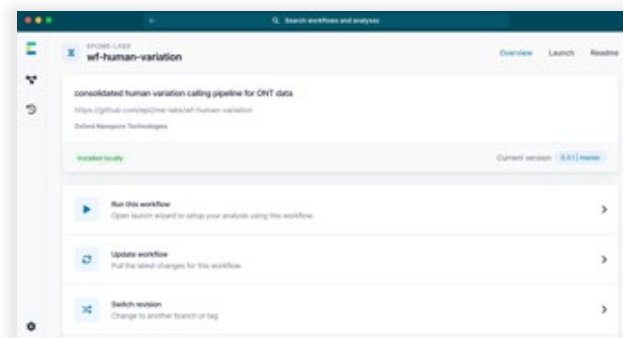


# Fully supported intuitive solutions

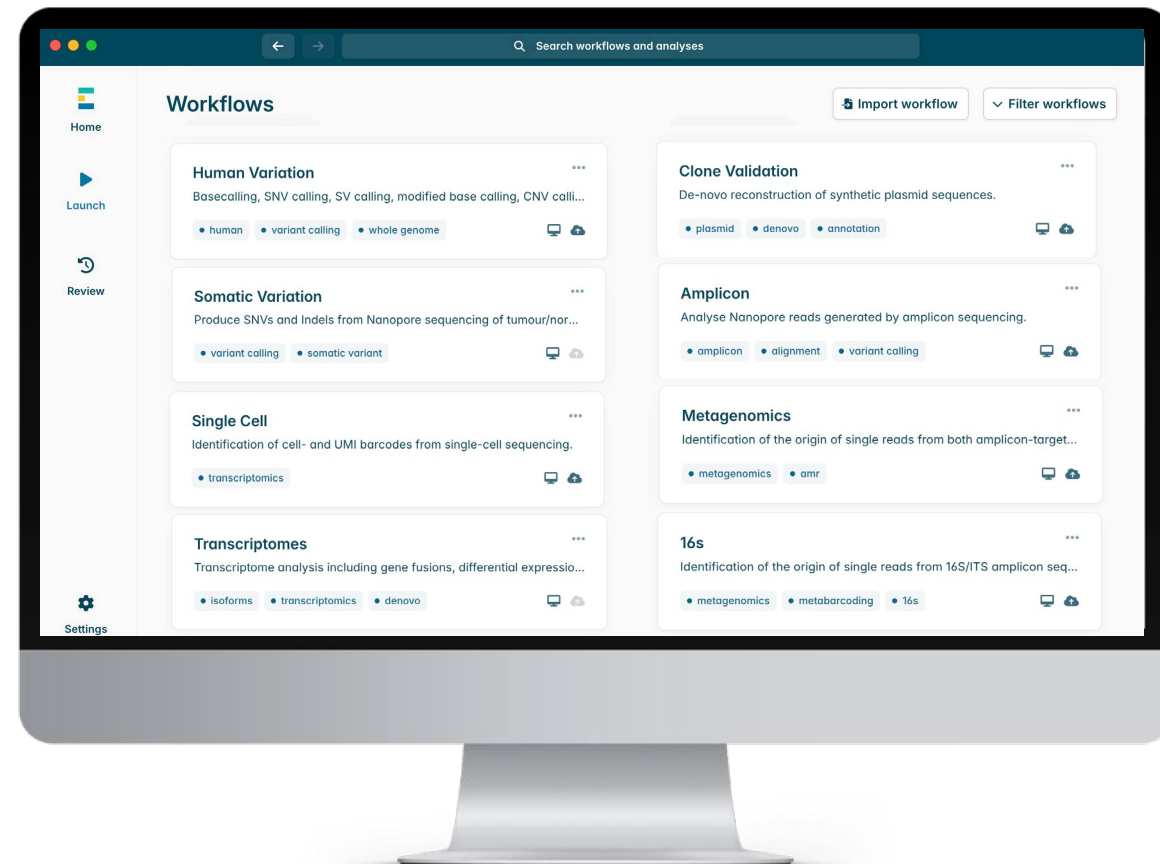
## Analyse your data with EPI2ME



- Intuitive interface
- Pre-configured workflows
- Intuitive, interactive reports
- Standard output files



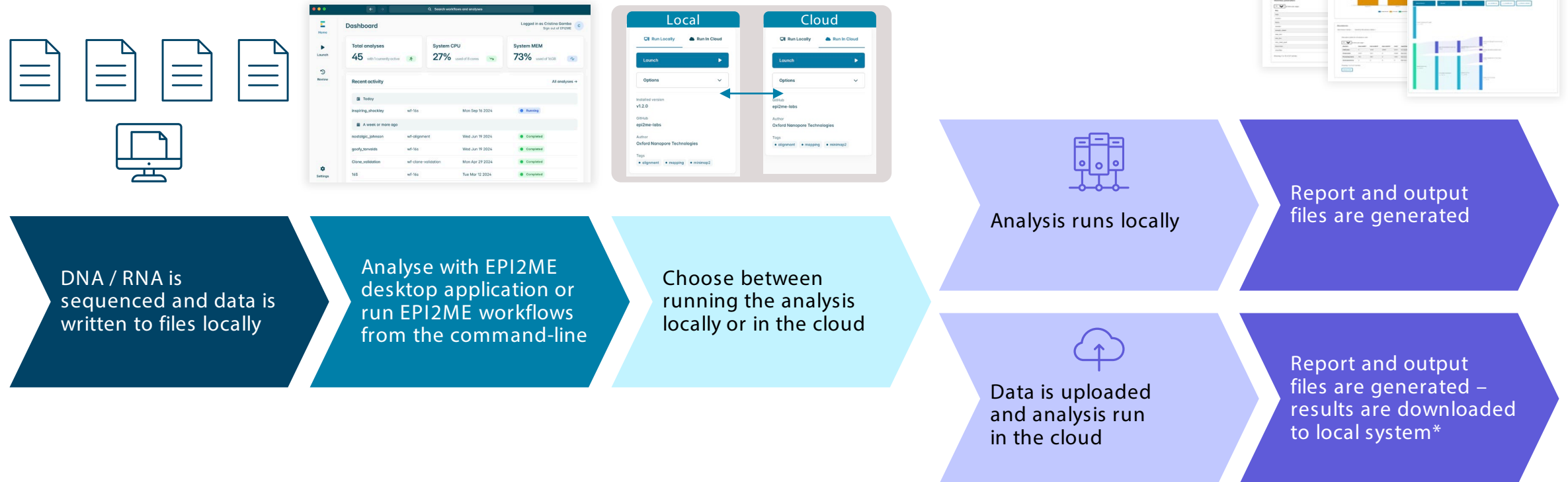
- Run locally or in the cloud
- Command-line access
- Easy integration





# EPI2ME: how does it work?

## Intuitive data analysis



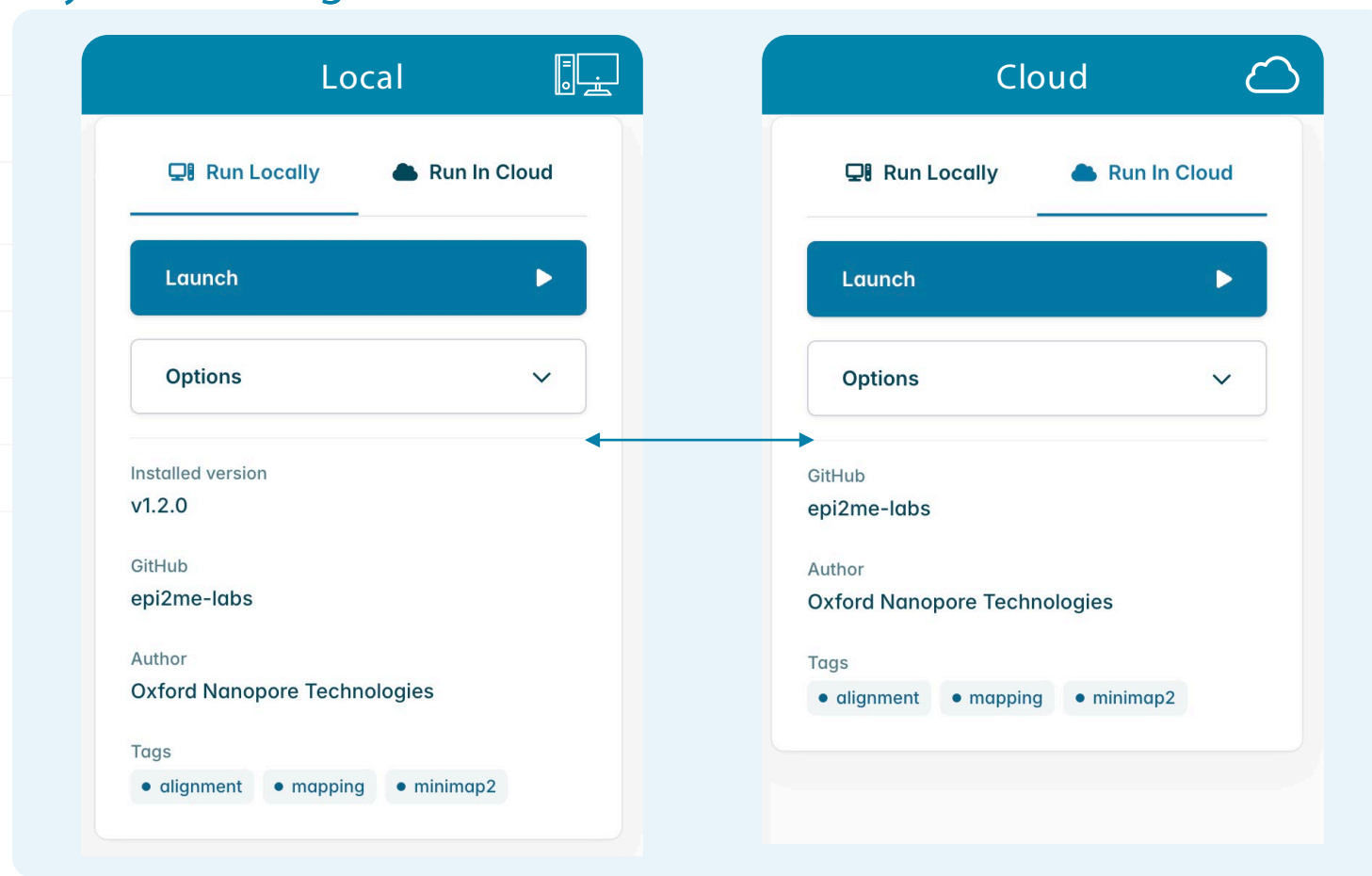
\* Data is not stored in the cloud and need to be synced to local systems. Data in the cloud is automatically removed after 2 weeks from the end of the analysis.



# Integrated local and cloud EPI2ME experience

Local and cloud analysis in a single software

|                          |
|--------------------------|
| Location                 |
| Access                   |
| Configurability          |
| Reporting                |
| Focus                    |
| Operating systems        |
| Bioinformatics expertise |





# EPI2ME desktop application: how does it work?

## An open analysis platform

➤ Free to download standalone **desktop application**

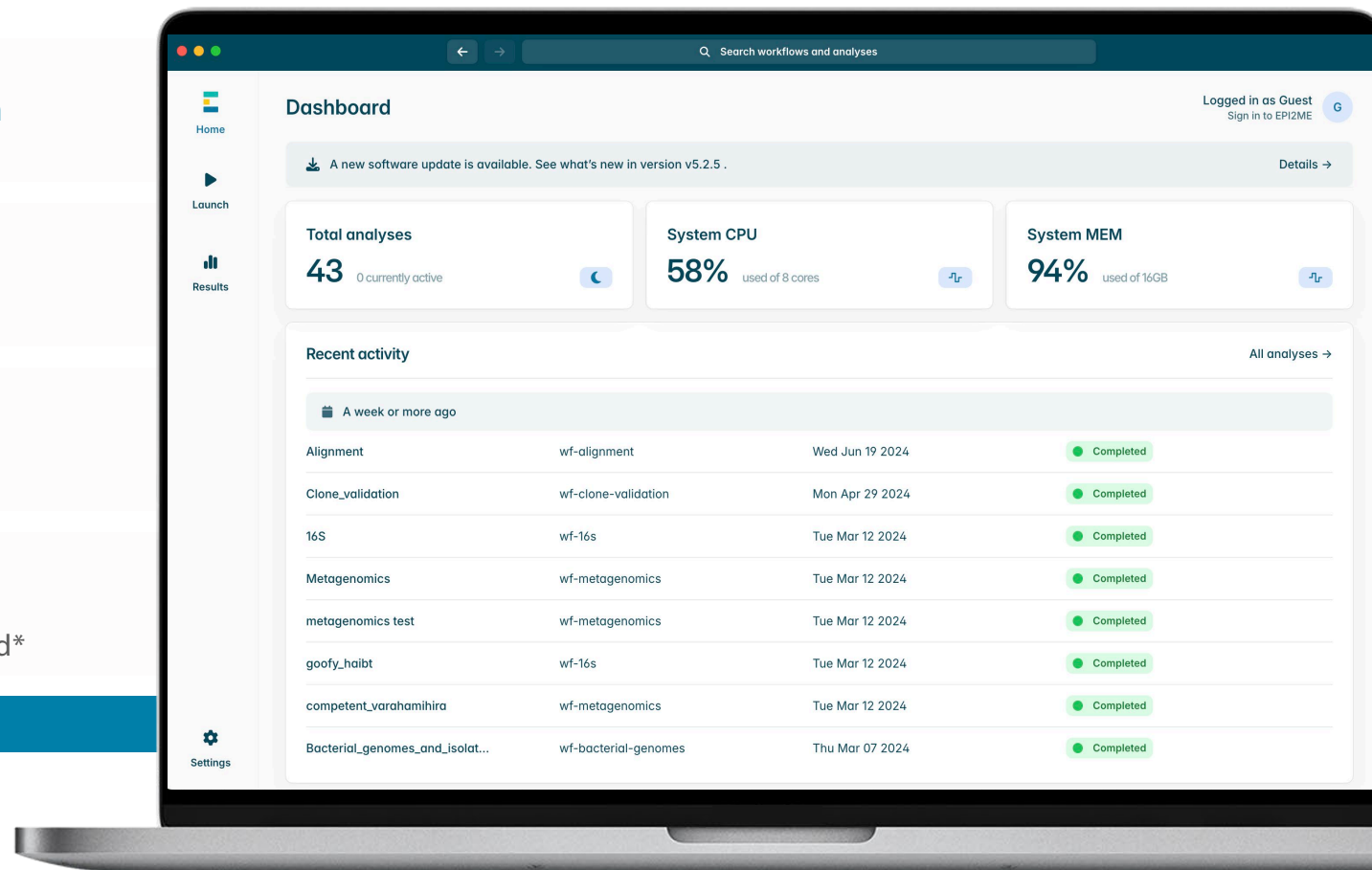
➤ Pre-packaged with **>15 workflows**

➤ **Minimal setup** required

➤ **Compatibility**


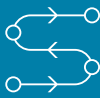

















- **Windows**, macOS, Linux
- Desktop computers, laptops, on-device, cluster or cloud\*


\*Must be logged in to access cloud functionality





# Workflows available in EPI2ME

|   |   |  |  |
|---|---|--|--|
|  <h2>Human genetics</h2> <div><div>wf-trio ●●● <span>NEW!</span></div><div>wf-human-variation ●●●</div><div>wf-somatic-variation ●●●</div></div>   |  <h2>Basic tasks</h2> <div><div>wf-alignment ●● </div><div>wf-basecalling </div></div> |  <h2>What's in my pot?</h2> <div><div>wf-metagenomics ●● </div><div>wf-16S ● </div></div>         |  <h2>Quality control</h2> <div><div>wf-clone-validation ● </div><div>wf-aav-qc ● </div></div> |
|  <h2>Transcriptome sequencing</h2> <div><div>wf-transcriptomes ●● </div><div>wf-single-cell ●●● </div></div> |  <h2>Targeted sequencing</h2> <div><div>wf-amplicon ● </div></div>  |  <h2>Infectious disease</h2> <div><div>wf-bacterial-genomes ●● </div><div>wf-flu ● </div></div> |  <h2>Genome structure</h2> <div><div>wf-teloseq ●● <span>NEW!</span></div><div>wf-pore-c ●●●</div></div>  |

 = cloud analysis available

 = large GPU required

### CPU requirements

- standard laptop 8 GB RAM -- Grid / P2i / P24 (/ Laptop)
- large memory computer 32 GB RAM -- Grid / P2i / P24 (/ Laptop)
- 16 CPUs + 64GB RAM required --- Prom24



Interested in running **wf-human-variation**  
in the cloud? Register your interest here





# EPI2ME workflows from command line

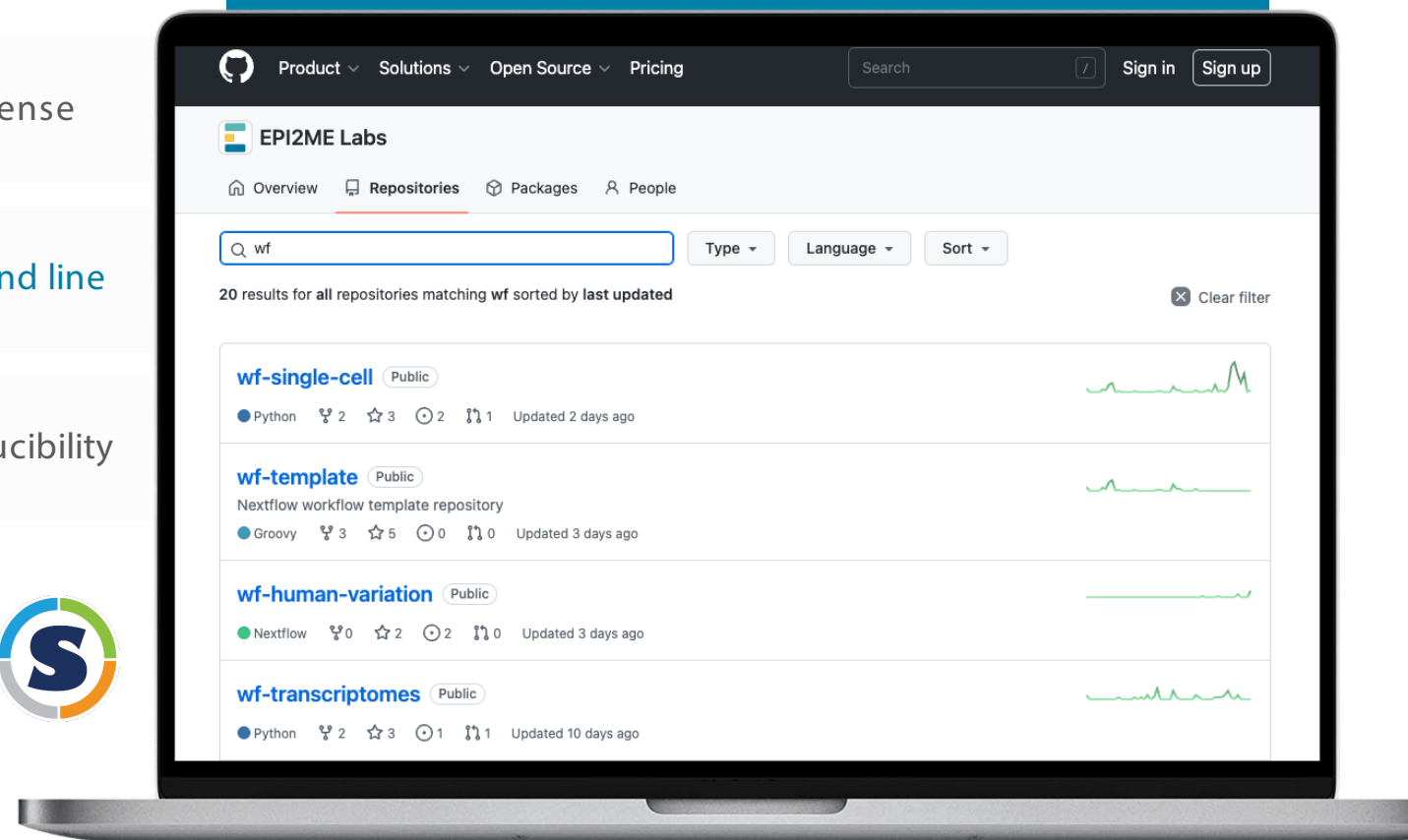
## For deployment on own clusters or custom pipelines

- EPI2ME code is deposited in [GitHub](#) under MPL-2.0 license
- Nextflow workflows can be executed from [the command line](#)
- Support containers Singularity and Docker for reproducibility



More at:  
[github.com/orgs/epi2me-labs/repositories](https://github.com/orgs/epi2me-labs/repositories)

Nextflow workflows available on GitHub **nextflow**





# EPI2ME integrations





# Workflows now easier to install

Enabled by the 2ME format



Supports offline  
compute  
infrastructure



2ME format  
includes all  
workflow pre-  
requisites



One-click  
installation



Workflow bundle

Code

Containers

Data

Reference data

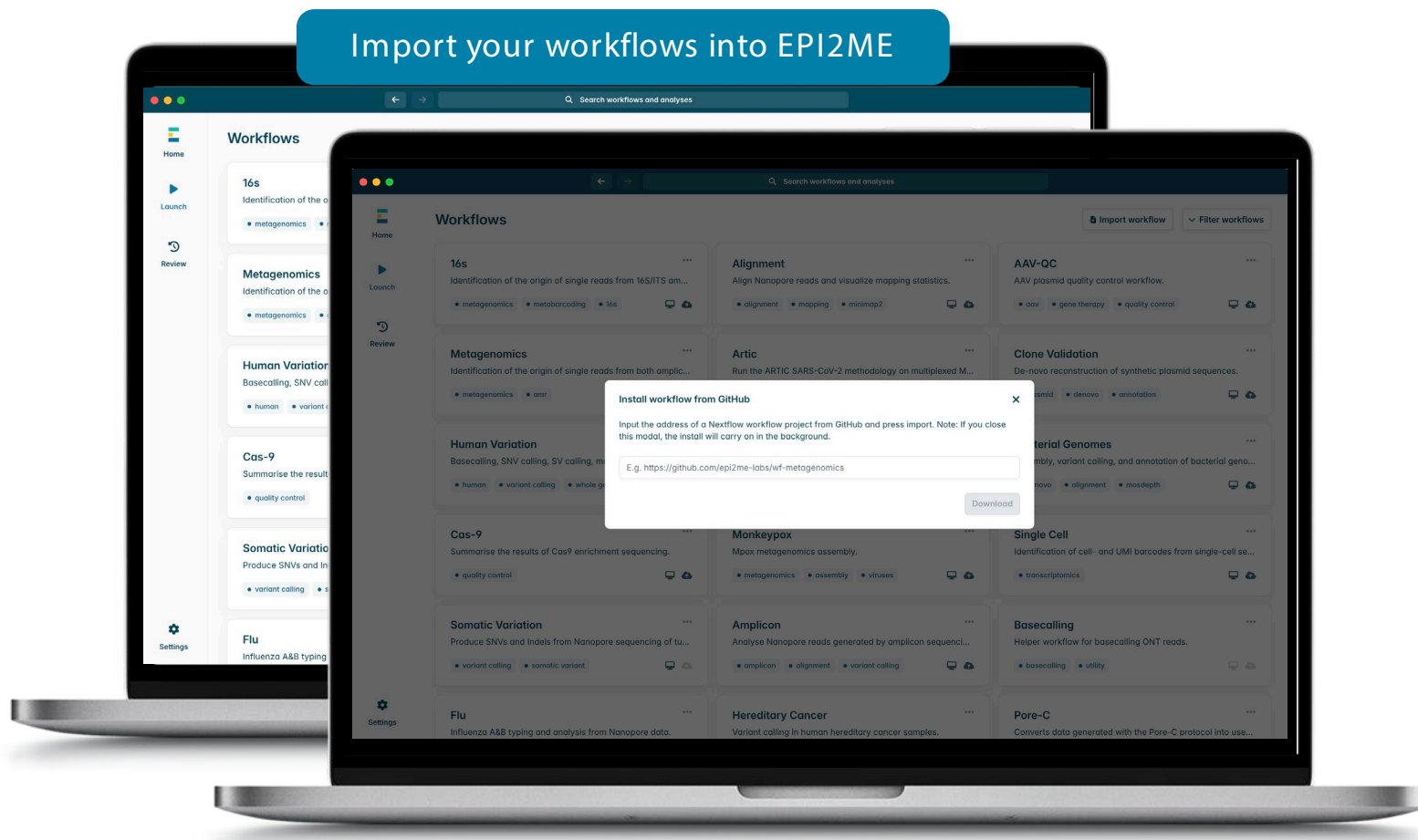
Instances



# Integrate your workflows with EPI2ME desktop application

## Install with the GitHub repository

- EPI2ME desktop application enables
  - Running locally community-developed bioinformatics workflows
  - Implemented in Nextflow as per nf-core standards
- Import straight from GitHub
- Share your workflows with the broader Oxford Nanopore community





# Integrate your workflows with EPI2ME desktop application

## Install with the GitHub repository

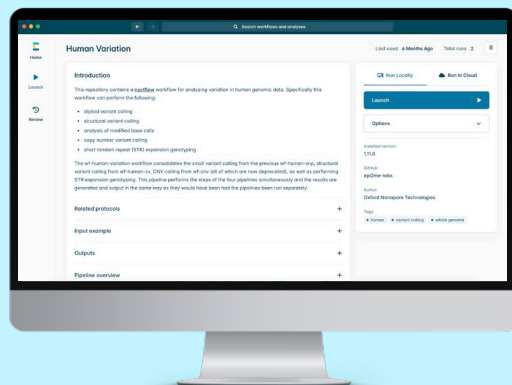
The screenshot displays the EPI2ME desktop application interface. On the left, a sidebar contains navigation links: Home, Launch, Results, and Settings. The main area is titled 'Workflows' and features a search bar at the top. Below the search bar, there are two buttons: 'Import workflow' (highlighted with an orange border) and 'Filter workflows'. A dropdown menu is open under 'Import workflow', showing two options: 'Import from Github' and 'Import a 2ME file'. The 'Import from Github' option is selected. Below the dropdown, there are three workflow cards: 'Mpox' (Mpox metagenomics assembly from nanopore sequencin...), 'Somatic Variation' (Human (somatic) SNV, SV, and modified base calling.), and 'Teloseq' (Analysis of telomeric sequence data generated with the ...). Each card has a list of tags and icons for launch and download. A modal window titled 'Install workflow from GitHub' is overlaid on the right side of the screen. It contains the following text: 'Input the address of a Nextflow workflow project from GitHub and press import. Note: If you close this modal, the install will carry on in the background.' Below the text is a text input field with the example URL 'E.g. https://github.com/epi2me-labs/wf-metagenomics'. At the bottom right of the modal is a 'Download' button. The modal also has a close button (X) in the top right corner. At the bottom of the main interface, there are 'Previous' and 'Next' navigation buttons.



# Enabling nanopore data interpretation

## Through tertiary analysis compatible partners

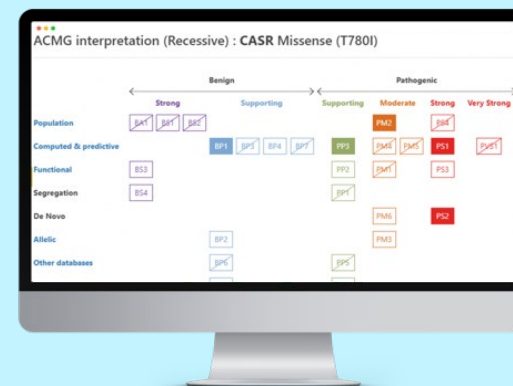
### EPI2ME Human variation workflow



Secondary analysis



### Compatible partners



Tertiary analysis and  
interpretation



# Datasets



# An increasing library of datasets

## > Oxford Nanopore open dataset project

- Test EPI2ME
- Tool development
- Reproducible benchmarks
- Explore characteristics of nanopore data

## > Over 20 datasets available e.g.

- Genome in a Bottle
- Plasmid Validation
- Metagenomics
- And much more



<https://epi2me.nanoporetech.com/dataindex/>

The screenshot shows the EPI2ME website interface. At the top, there's a navigation bar with 'Resources' and 'About' dropdowns, a search bar, and a dark mode toggle. The main content area is titled 'Latest datasets and analyses' and includes a brief description. Below this is a section for 'Human genomics' containing a table of datasets. To the right, there's a 'Table Of Contents' sidebar with a list of links.

| Dataset  | Flow Cell   | Kit        | Basecall model       | EPI2ME workflows                     | Date       |
|--|-------------|------------|----------------------|--------------------------------------|------------|
| Genome in a Bottle Data Release 2025.01                            | FLO-PRO114  | SQK-LSK114 | v5.0.0 SUP and HAC   | wf-basecalling, wf-human-variation   | 2025-01-26 |
| Modified Base Best Practices and Benchmarking                      | FLO-PRO114  | SQK-LSK114 | v5.0.0 SUP and HAC   |                                      | 2024-10-22 |
| Nanopore-only T2T assembly of a human genome                       | FLO-PRO114  | SQK-LSK114 | v4.3.0 SUP and HAC   |                                      | 2024-05-22 |
| Telomere sequencing  | FLO-PRO114M | SQK-LSK114 | v5.0.0 SUP and HAC   | wf-teloseq                           | 2024-05-21 |
| Updated Tumor Normal Pair Benchmark Dataset                        | FLO-PRO114  | SQK-LSK114 | v4.2.0 SUP and HAC   | wf-basecalling, wf-somatic-variation | 2024-03-07 |
| An experimental extremely high-accuracy, ultra-long sequencing kit | FLO-PRO114  | SQK-ULK114 | Bespoke dorado model |                                      | 2023-12-06 |
| Reduced Representation Methylation Sequencing (RRMS)               | FLO-MIN106  | SQK-LSK110 | v3.3                 |                                      | 2022-07-27 |

**Table Of Contents**

- 1 Latest datasets and analyses
- 2 Human genomics
- 3 Single-cell, transcriptomics and direct RNA
- 4 Synthetic Biology
- 5 Plant and Animal
- 6 Metagenomics
- 7 Use of Oxford Nanopore Open Data
- 8 Archived datasets
- 9 Data access





# EPI2ME workflows

- Examples of streamlined analysis workflows
- Basecalling and mapping to a reference



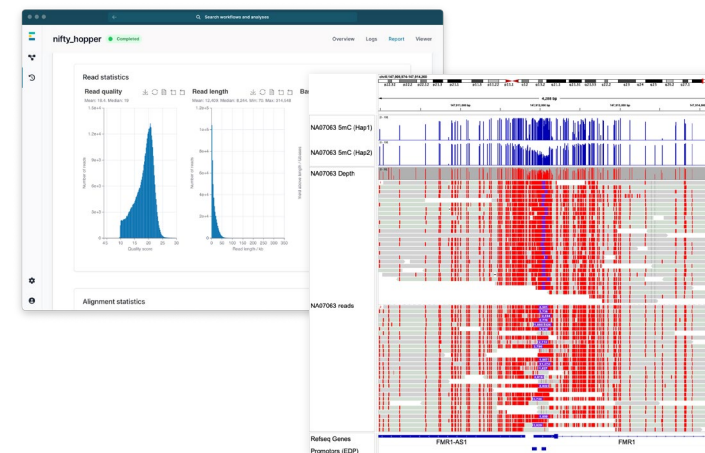
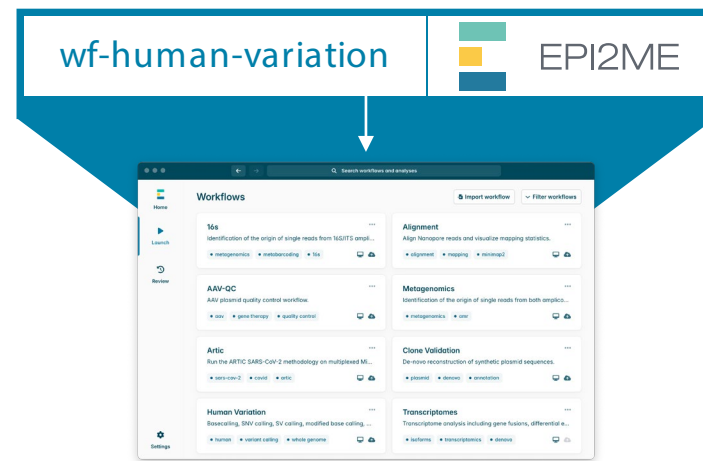
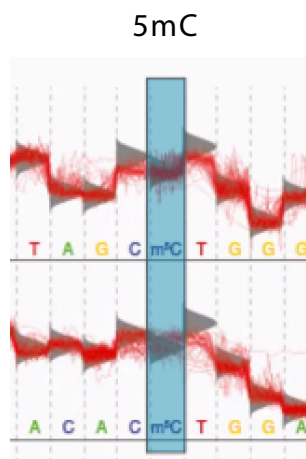
# Human genetics





# Investigate human variation with a single workflow

Investigation of structural variants, single nucleotide variants, and methylation using EPI2ME



Sample sequencing and  
basecalling in MinKNOW

wf-human-variation

Detect  
SNVs, SVs, STR, CNVs,  
methylation, phasing

Intuitive reports  
Visualise results

MinKNOW



This workflow is accessible from both the intuitive graphical interface and the command line

© 2025 Oxford Nanopore Technologies plc. All rights reserved.  
Oxford Nanopore Technologies products are not intended for use for health  
assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

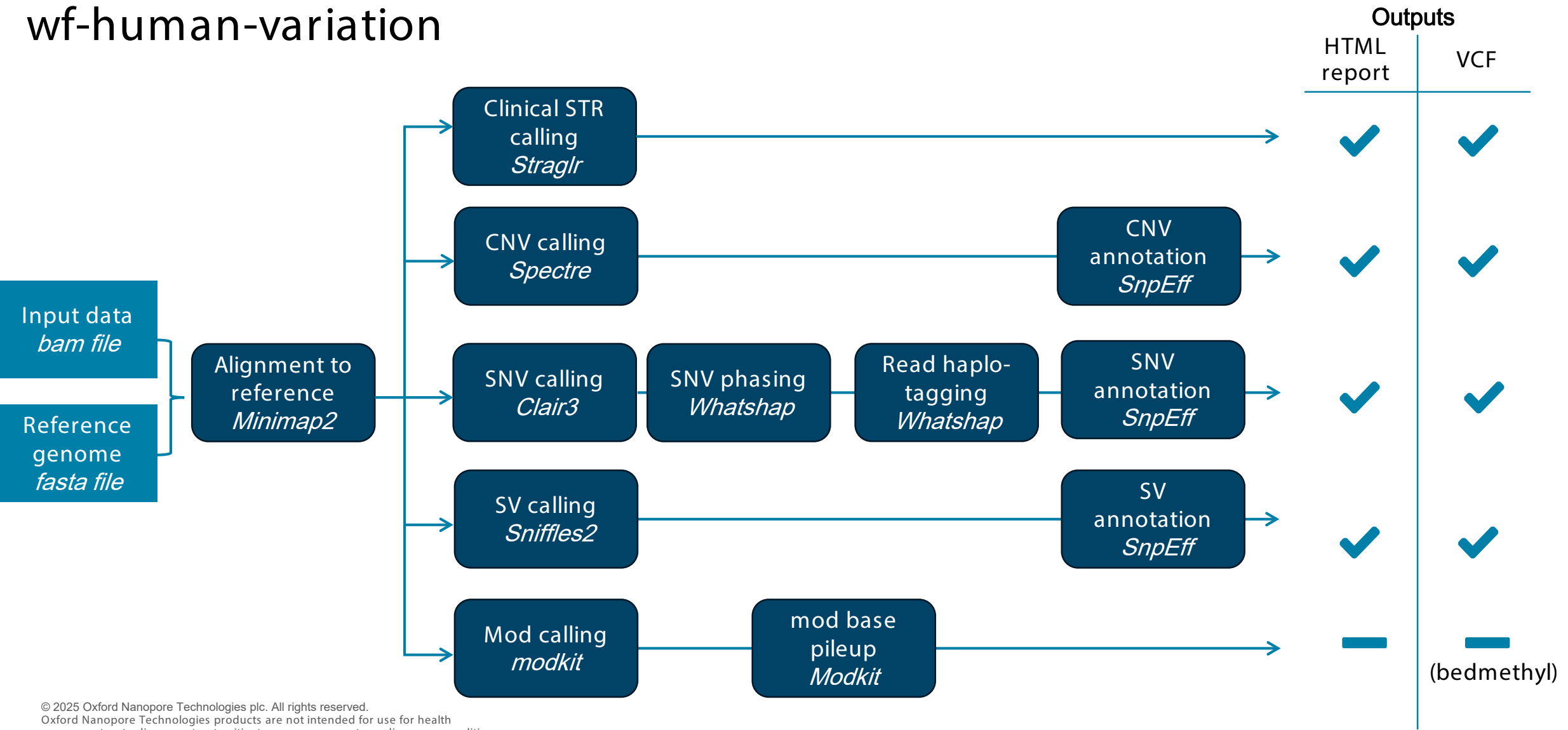


Interested in running **wf-human-variation**  
in the cloud? Register your interest here



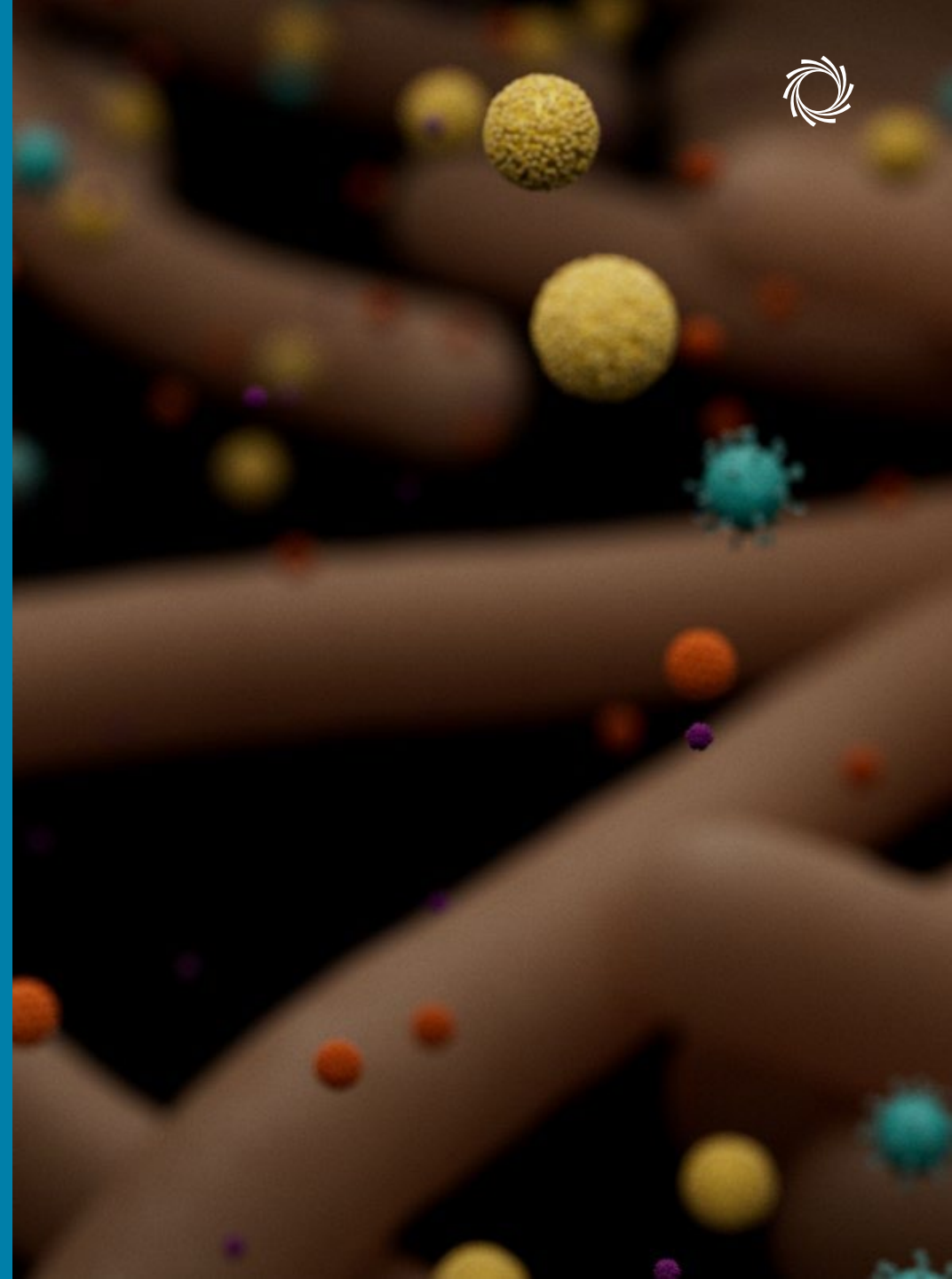


# wf-human-variation





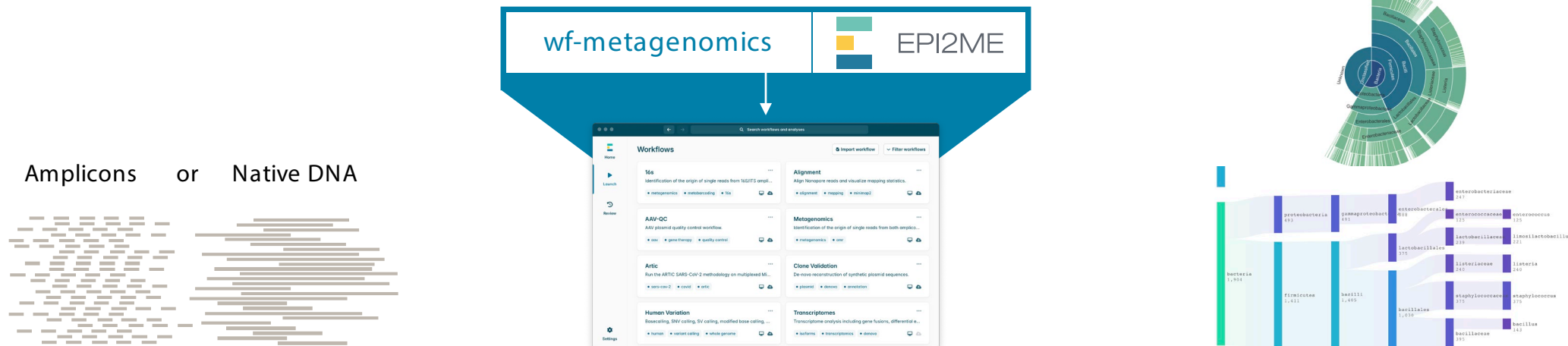
# What's in my pot?





# Get insights into your metagenomic samples: shotgun or amplicons

Intuitive visualisation of taxonomy, diversity, abundances, and more



Sequence and basecall  
amplicon-targeted and  
shotgun metagenomics  
samples in MinKNOW

wf-metagenomics

Quick classification with  
Kraken 2. Fine classification  
with minimap2  
Real-time options

Intuitive report, classified  
and unclassified reads,  
and text files with  
lineage details

MinKNOW

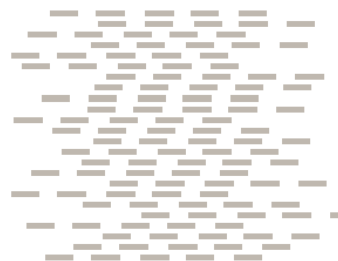


This workflow is accessible from both the intuitive graphical interface and the command line

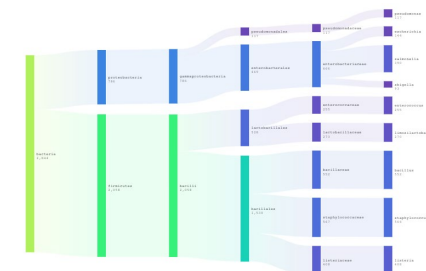


# Identification based on 16S, 18S, and ITS amplicons

Intuitive visualisation of taxonomy, diversity, abundances, and more



16S / 18S / ITS Amplicons



Sequence and basecall  
amplicon-targeted  
16S/18S/ITS in MinKNOW

wf-16S

Quick classification with  
Kraken 2. Fine classification  
with minimap2  
Real-time options

Intuitive report, classified  
and unclassified reads,  
and text files with  
lineage details

**MinKNOW**



This workflow is accessible from both the intuitive graphical interface and the command line



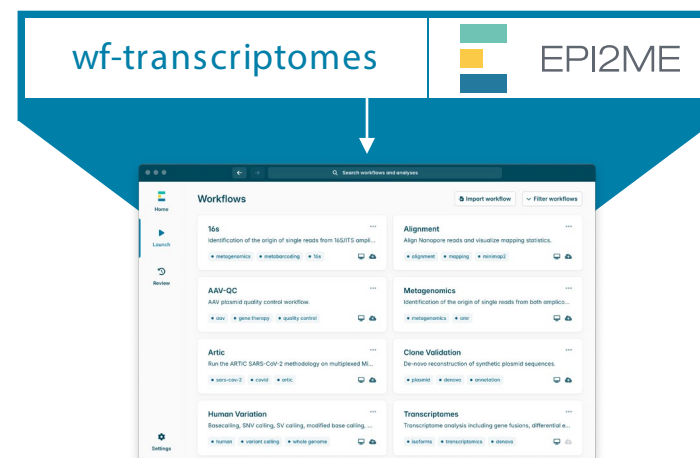
# Transcriptome sequencing



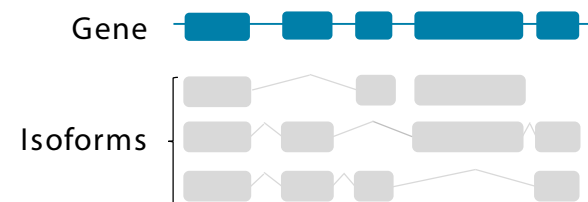


# Transcriptomics workflow for cDNA or native RNA

Identify isoforms, measure gene expression, and detect fusion genes



Native RNA or  
cDNA reads



Full-length native RNA  
reads or cDNA transcripts  
sequenced and  
basecalled in MinKNOW

wf-transcriptomics

Input FASTQ file(s) and  
reference genome (if available).  
Additional reference files  
needed for gene fusion and  
gene expression\*

Assembled transcriptome  
(annotated). Optional: fusion  
transcript sequences and  
differential transcript plots

MinKNOW

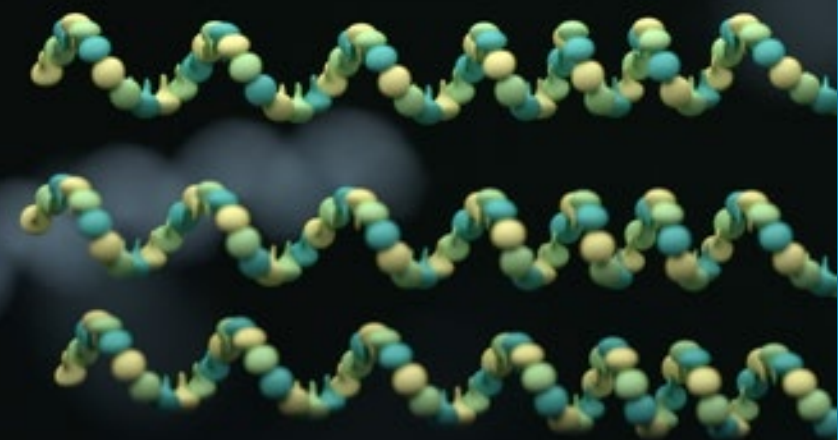
EPI2ME

\*For differential expression analysis provide reference annotation in GFF2/3 format, and for fusion detection, provide the JAFFAL reference files. This workflow is accessible from both the intuitive graphical interface and the command line.





# Targeted sequencing







# Get high-quality amplicon consensus & detect variants

Barcode and pool samples & targets



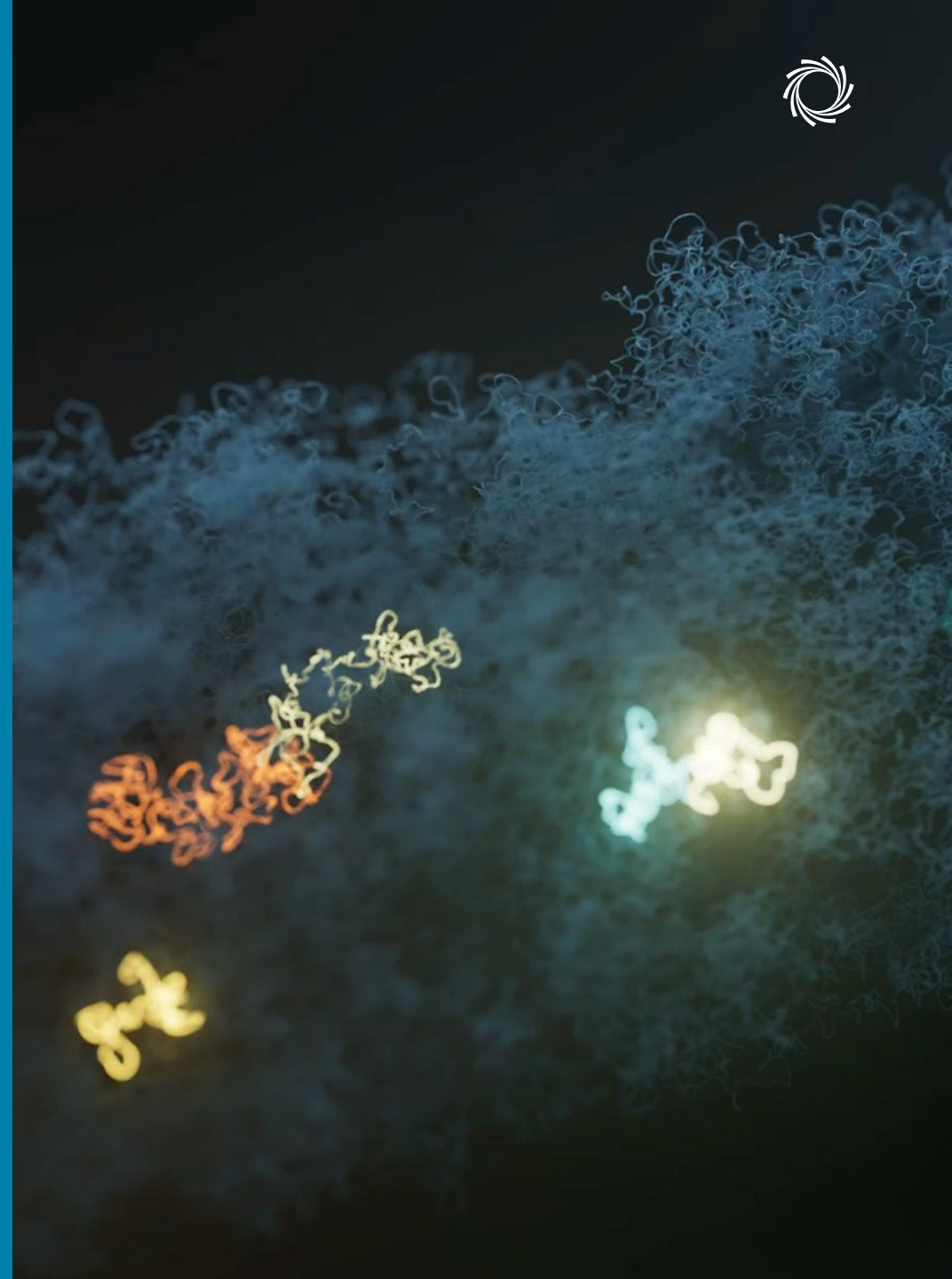
MinKNOW



This workflow is accessible from both the intuitive graphical interface and the command line



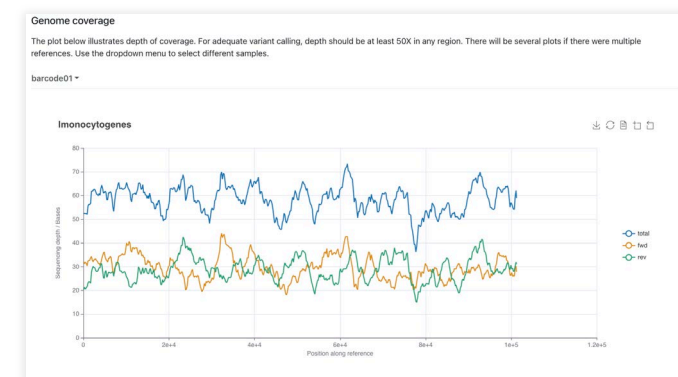
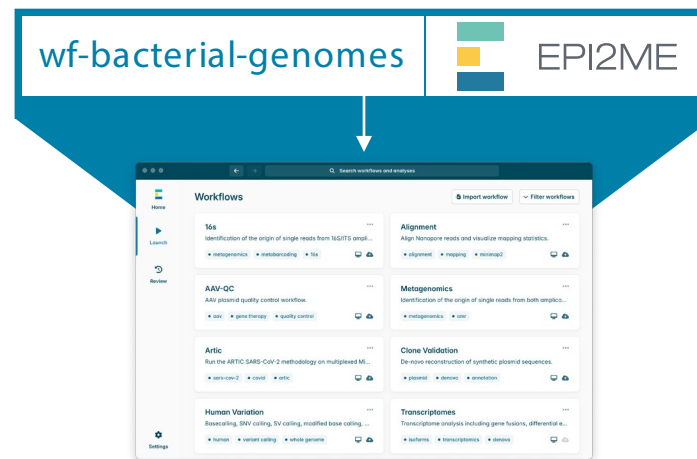
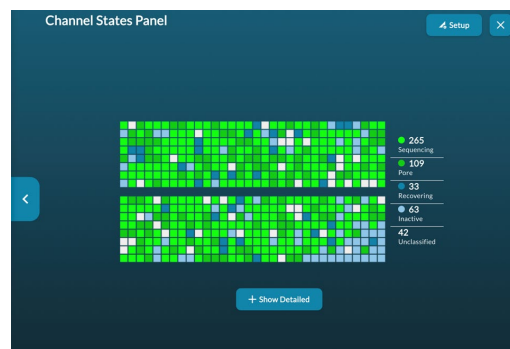
# Infectious disease research





# Bacterial genomes alignment or assembly

## Including bacterial annotation & antimicrobial resistance genes



Sample sequencing and basecalling in MinKNOW

wf-bacterial-genomes

Alignment to reference or assembly. Consensus annotation and resistance genes

Intuitive report, FASTA consensus and annotation

MinKNOW



This workflow is accessible from both the intuitive graphical interface and the command line



# Emerging tools for advanced users

- Open access software
- Bioinformatics expertise needed



# The latest emerging tools from Oxford Nanopore

## Open access software on GitHub

The latest tools and algorithms developed by Oxford Nanopore are available on the Oxford Nanopore GitHub repository

- From basecallers to modified base analysis
- Ready for your custom pipelines
- Command-line experience required
- Limited support due to rapid evolution

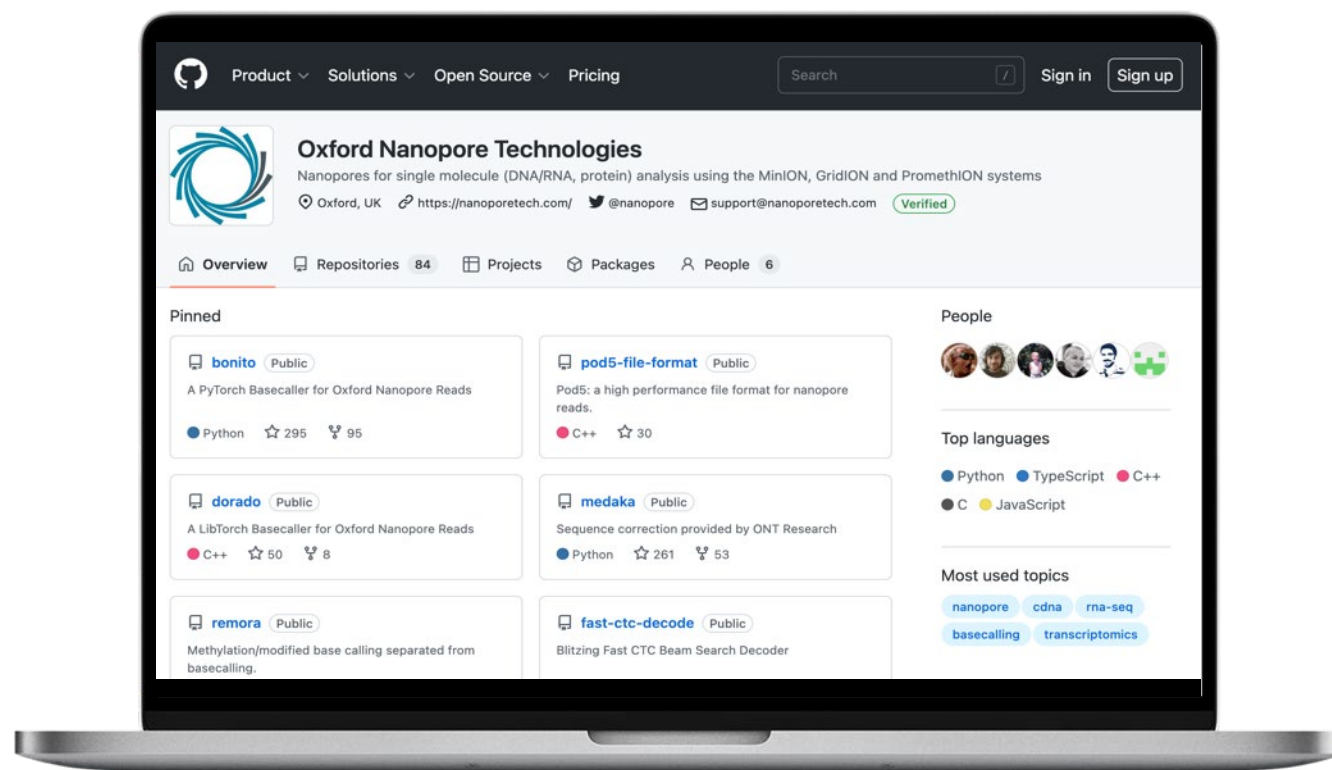
### Key tools available

- Dorado: latest basecaller and other tools available standalone
- Modkit: toolbox for working with modified bases



More at: [github.com/nanoporetech](https://github.com/nanoporetech)

### Oxford Nanopore GitHub repository

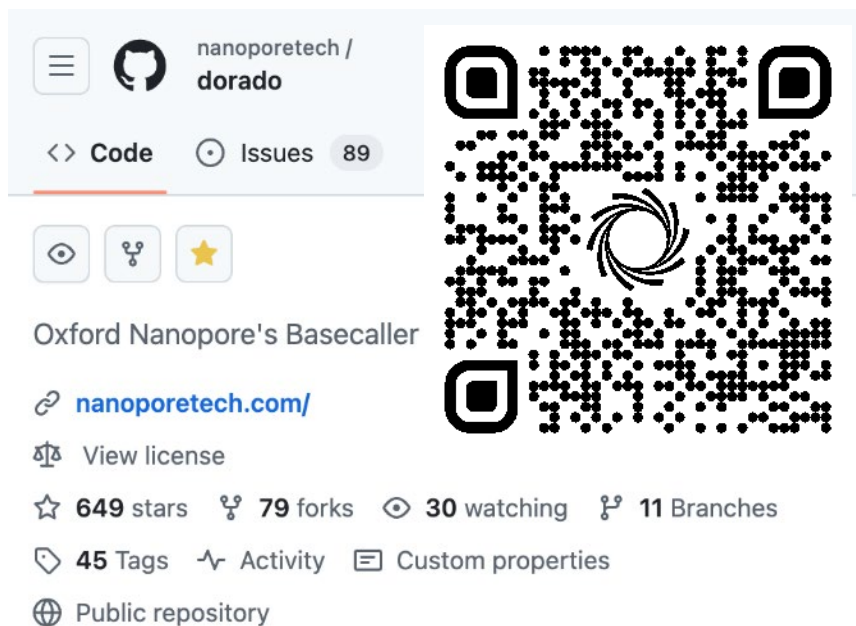




# What is Dorado?

Dorado powers Oxford Nanopore basecalling and more

## Latest features available on GitHub

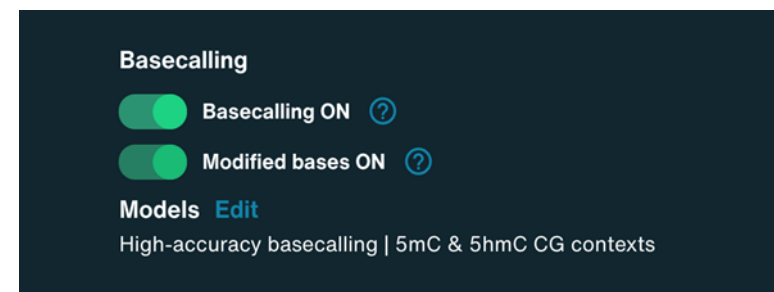


The screenshot shows the GitHub repository for 'nanoporetech / dorado'. It includes the repository name, a QR code, and various statistics: 649 stars, 79 forks, 30 watching, 11 branches, and 45 tags. The repository is described as 'Oxford Nanopore's Basecaller' and is a 'Public repository'.

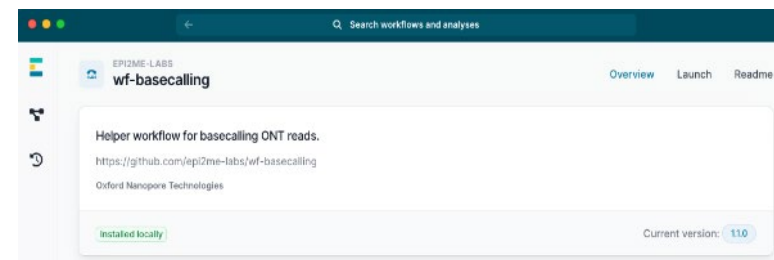


## Integrated in MinKNOW and EPI2ME

### MinKNOW



The MinKNOW interface shows basecalling settings. 'Basecalling' is turned ON, and 'Modified bases' are also ON. The 'Models' section is highlighted in blue, with a link to 'Edit'. Below this, it states 'High-accuracy basecalling | 5mC & 5hmC CG contexts'.

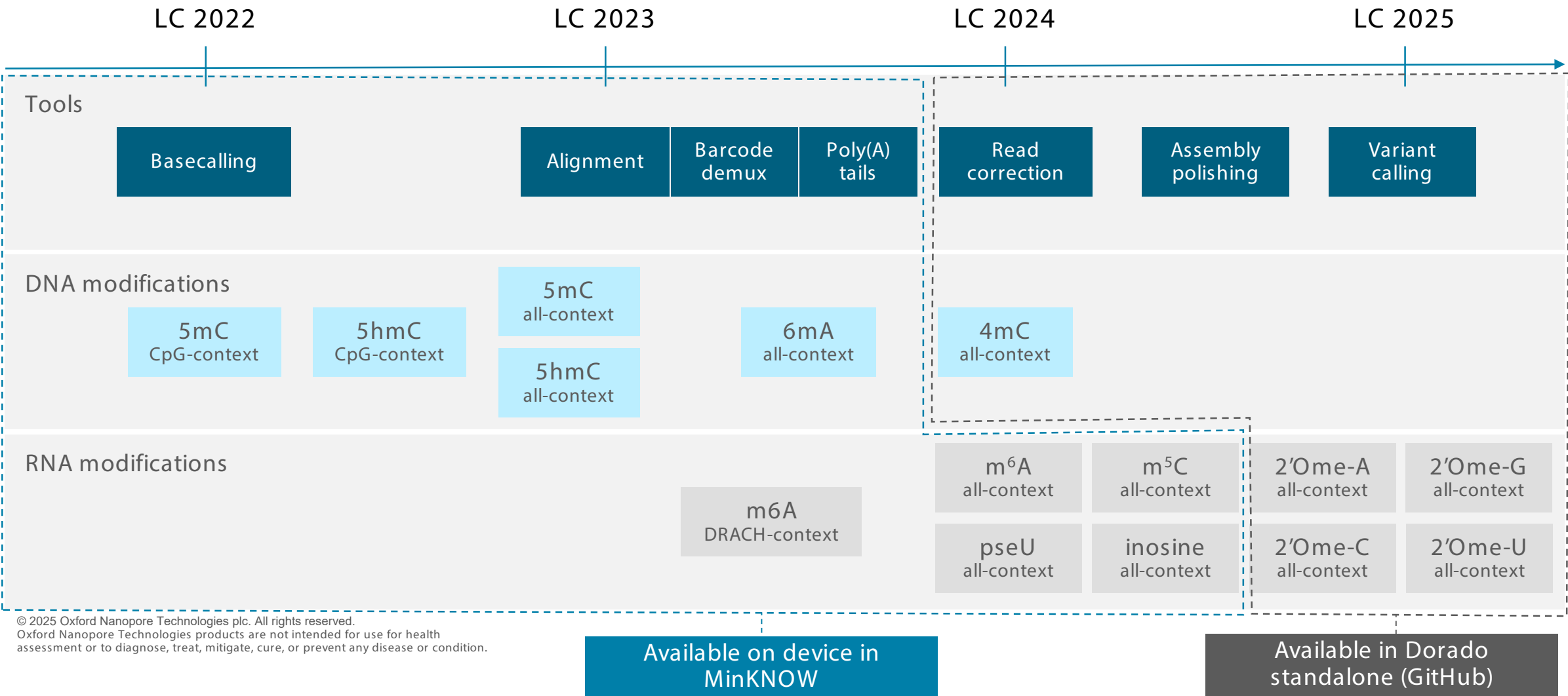


The EPI2ME interface shows a workflow page for 'wf-basecalling'. It includes a description: 'Helper workflow for basecalling ONT reads.' and a link to the GitHub repository. The status 'Installed locally' is shown, along with the 'Current version: 1.1.0'.



# Dorado: the engine for Oxford Nanopore data analysis

From basecalling through to assembly polishing and small variant calling







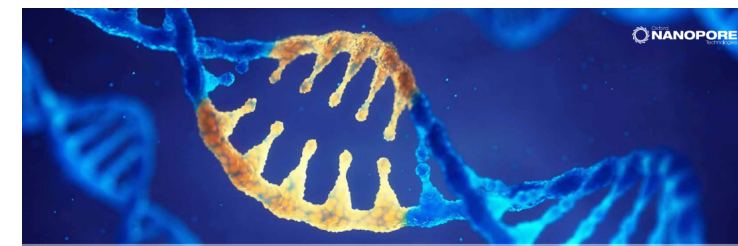
# Methylation analysis with Modkit

Available in GitHub – implemented in EPI2ME workflows

- Modkit is a suite of tools for manipulating modified-base data
- Integrated in EPI2ME wf-human-variation
- Best-practices filtering implemented
- Suitable for the analysis of 5mC and 5hmC
- Compatible with chromatin stenciling (6mA): find and qualify promoter regions
- Compatible with hemi-methylation
- Input: BAM files (with MM and ML tags)
- Output: summary counts of modified and unmodified bases (bedMethyl)



More at: [github.com/nanoporetech/modkit](https://github.com/nanoporetech/modkit)



## Accessible and robust base modification analysis with Modkit, the multi-tool for nanopore epigenetics

Leveraging DNA and RNA methylation data in your experiment shouldn't be hard – so we made a tool to make it easy. Modkit is open source and integrated into EPI2ME™ workflows

Contact: [art.rand@nanoporetech.com](mailto:art.rand@nanoporetech.com) More information at: [nanoporetech.github.io/modkit/](https://nanoporetech.github.io/modkit/)

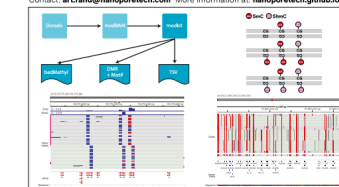


Fig. 1 Genome browser showing methylation tracks generated by Modkit

### The first step in base modification steps: 'pileup' and 'extract'

Modified base detection is integrated into Oxford Nanopore basecalling software Dorado and MinKNOW™. The first step is to aggregate base modification counts across genomic or transposonic positions. This can be done by performing the 'pileup' command, or the duplex-enabled count 'pileup-hem'. These tables, bedMethyl tables, are the input to many downstream steps and can be parsed with common analysis software. The 'extract' command will generate per-read tables allowing deeper investigation into single-molecule base modifications. The 'pileup-hem' command will count the occurrences of double-stranded methylation patterns (hemi-methylation) when provided with duplex reads. Finally, stats will collapse genomic windows or regions into summary statistics.

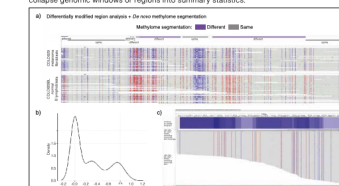


Fig. 3 Genome browser showing automatic detection of differentially methylated regions (DMRs), and high scoring CpG island

### Explore differentially methylated regions with automatic segmentation, using 'dmr' and 'entropy'

The Modkit tool suite contains flexible and intuitive differential methylation analysis algorithms that leverage multiple modification types (e.g. 5mC and 5hmC). A bedMethyl table (generated by 'pileup') along with regions of interest (for example, CpG islands) can be quickly scanned for differential methylation, ranked, and tagged. Emphasis has been placed on making the output intuitive and enabling exploratory data analysis. Alternatively, de novo segmentation can be performed, as shown for the COLO829 tumor-normal research samples (Fig. 3a); differentially methylated regions are highlighted in purple. Direct RNA sequencing of HEK2936 samples revealed a distribution of methylation entropy (Fig. 3b), with an example high-entropy transcript shown here (Fig. 3c).

Information content of genes or transcripts may be subject to change.

Oxford Nanopore Technologies, the Oxford Nanopore logo, EPI2ME, Dorado, MinKNOW, and PromethION are registered trademarks or the subject of trademark applications of Oxford Nanopore Technologies plc in various countries. Information contained herein may be protected by patents or other intellectual property of Oxford Nanopore Technologies plc. All other brands and names are the property of their respective owners. © 2024 Oxford Nanopore Technologies plc. All rights reserved. Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

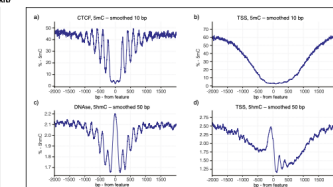


Fig. 2 Inspect modification patterns localised to genomic features of interest

### Visualise and compare methylation patterns across genomic annotations with 'localize'

Changes in base modification patterns can be indicative of biological mechanisms. The 'localize' command quickly aggregates base modification rates localised to genomic features of interest. The tabulation is fast, speeding up iteration and exploratory data analysis. Aggregation is performed per base modification, allowing inspection of, for example, 5mC and 5hmC in the same sample. Example plots show 5mC patterns at CTCF binding locations (Fig. 2a) and transcription start sites (TSS) (Fig. 2b). Peaks of 5hmC can be seen at DNase hypersensitivity sites (Fig. 2c) and TSS (Fig. 2d).

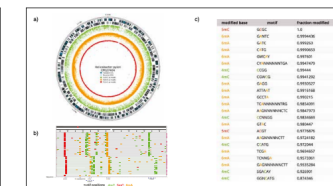


Fig. 4 Genome browser showing discovered methylation motifs and example output table

### Automatically find sequence motifs enriched for methylation with 'find-motifs'

Methyltransferase enzymes will methylate specific sequence motifs. Nanopore sequencing can detect the most common prokaryotic modifications (5mC/5hmC and 6mA) at the same time, allowing the discovery of adenine and cytosine methyltransferase motifs. In (Fig. 4a), we show a browser view of Helicobacter pylori sequencing reads with 5mC, 5hmC, and 6mA. Discovered motifs, in an example table from the same organism (Fig. 4b), 23 de novo discovered motifs are seen, with varying levels of modification and specificity.

Download poster at: <https://shorturl.at/oDIiY>





# Our user community continually releases analysis tools

## User-developed tools and algorithms tailored to Oxford Nanopore data

### Common tools

- E.g., demultiplexing, filtering, mapping, assembling, variant calling, etc.

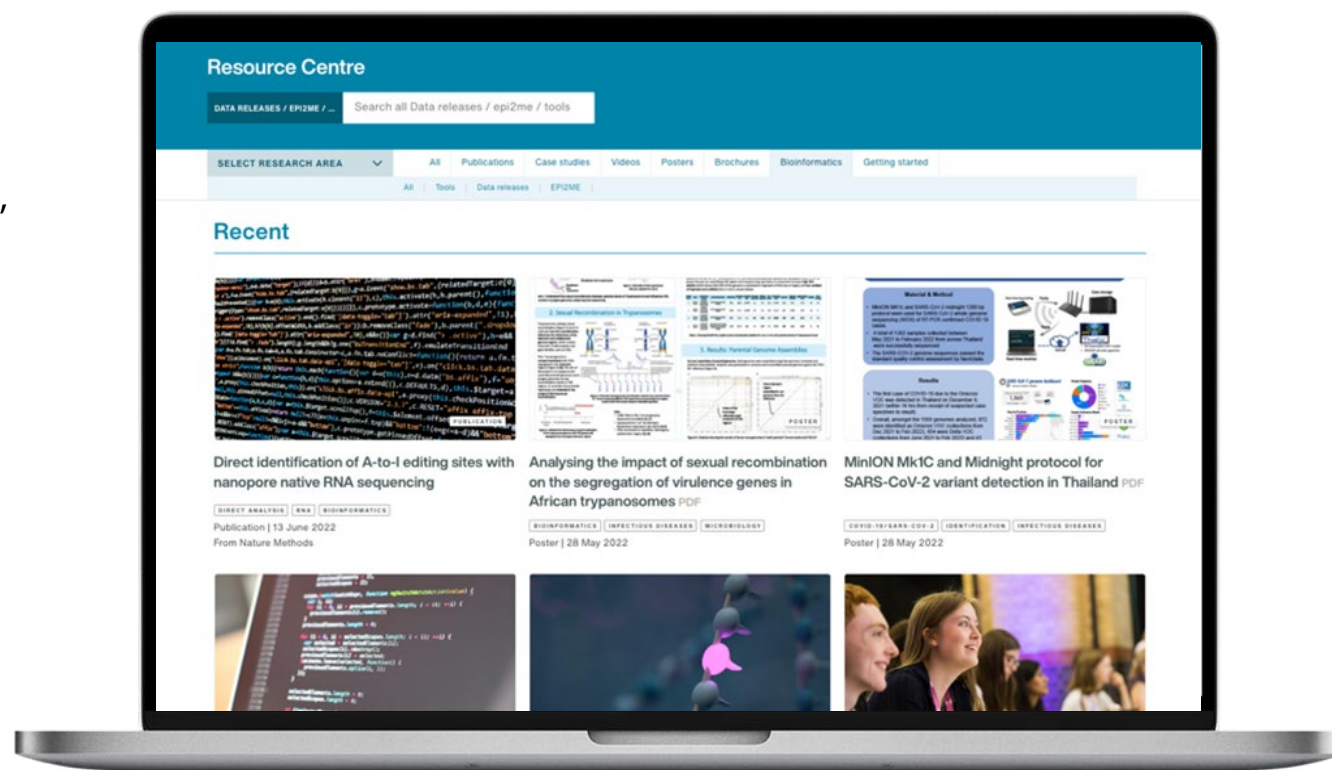
### Application-specific tools

- E.g., bacterial genome assembly, SARS-CoV-2 monitoring, hepatitis C genome sequencing, mitochondrial DNA haplogroup classification, novel pathogen detection, transposon insertion identification, etc.

### Extended functionality

- E.g., visualization tools for methylated data, RNA methylation prediction, barcode aware adaptive sampling

The bioinformatics section of the nanoporetech.com resource center



More at: [nanoporetech.com/resource-centre](https://nanoporetech.com/resource-centre)



# Basecalling with Dorado

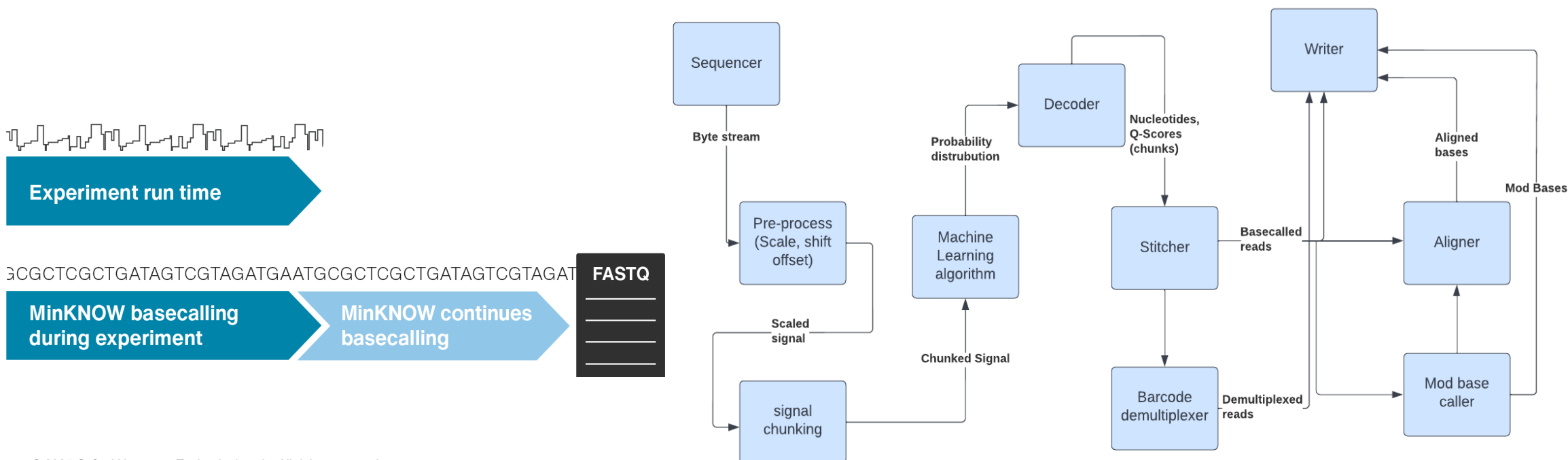


# Dorado – a new basecaller

Dorado is the successor to our previous basecaller known as Guppy

- Improvements to speed & usability (equivalent accuracy)

The basecaller uses machine learning models (e.g *dna\_r10.4.1\_e8.2\_400bps\_sup@v4.3.10*) to translate raw signal data (squiggles) into nucleotide sequence data

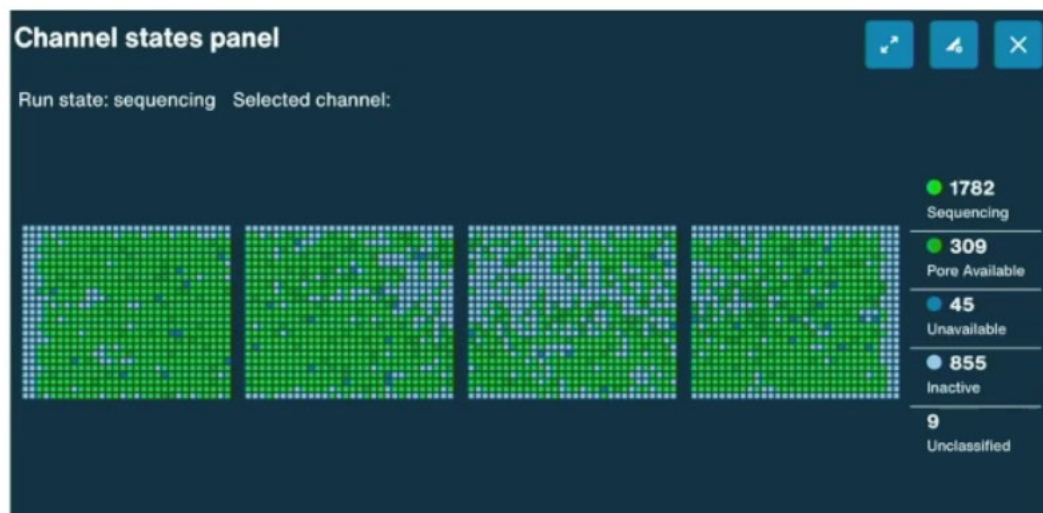




# Dorado – a new basecaller

operates as a stand-alone tool, and a MinKNOW-integrated basecaller

- Dorado is a stand -alone, command line application ( [www.github.com/nanoporetech/dorado](https://www.github.com/nanoporetech/dorado) )
  - Stand -alone version is usually a version ahead (and often access to newer models/feature -sets)
- Dorado is also integrated into MinKNOW for live -basecalling and in both MinKNOW and EPI2ME (Linux-only) for post -run basecalling.



```
steven.batinovic — zsh — 123x34

Last login: Fri Mar 22 22:34:45 on console
(base) steven.batinovic@X4NVGK4F41 ~ % dorado basecaller hac pod5/DNA/ > calls.bam
[2024-03-26 06:38:30.366] [info] Assuming cert location is /etc/ssl/cert.pem
[2024-03-26 06:38:30.367] [info] - downloading dna_r10.4.1_e8.2_400bps_hac@v4.3.0 with httpLib
[2024-03-26 06:38:31.306] [info] > Creating basecall pipeline
[2024-03-26 06:38:37.235] [info] - set batch size to 480
[2024-03-26 06:39:55.330] [info] > Simplex reads basecalled: 4000
[2024-03-26 06:39:55.330] [info] > Basecalled @ Samples/s: 1.059702e+06
[2024-03-26 06:39:55.394] [info] > Finished
(base) steven.batinovic@X4NVGK4F41 ~ %
```



# Decoding basecaller model names

| Basecalling Models                  | Compatible Modifications                 | Modifications Model Version | Data Sampling Frequency |
|-------------------------------------|--|-----------------------------|-------------------------|
| dna_r10.4.1_e8.2_400bps_fast@v5.2.0 |  |                             | 5 kHz                   |
| dna_r10.4.1_e8.2_400bps_hac@v5.2.0  | 4mC_5mC<br>5mCG_5hmCG<br>5mC_5hmC<br>6mA | v1<br>v1<br>v1<br>v1        | 5 kHz                   |
| dna_r10.4.1_e8.2_400bps_sup@v5.2.0  | 4mC_5mC<br>5mCG_5hmCG<br>5mC_5hmC<br>6mA | v1<br>v1<br>v1<br>v1        | 5 kHz                   |



# Decoding basecaller model names

- Dorado models names are systematically structured. Example: dna\_r10.4.1\_e8.2\_400bps\_sup@v4.3.0
- **Analyte Type (dna):** denotes the type of analyte being sequenced. For DNA sequencing, it is represented as dna. If you are using the Direct RNA Sequencing Kit, this will be rna.
- **Pore Type (r10.4.1):** This section corresponds to the type of flow cell used. For instance, FLO-MIN114/FLO-FLG114 is indicated by r10.4.1, while FLO-MIN106D/FLO-FLG001 is signified by r9.4.1.
- **Chemistry Type (e.8.2):** This represents the chemistry type, which corresponds to the kit used for sequencing. For example, Kit 14 chemistry is denoted by e.8.2.
- **Translocation Speed (400bps):** speed of translocation.
- **Model Type (sup):** This represents the size of the model, where larger models yield more accurate basecalls but take more time. The three types of models are fast, hac, and sup. The fast model is the quickest, sup is the most accurate, and hac provides a balance between speed and accuracy. For most users, the hac model is recommended.
- **Model Version Number (v4.3.0):** This denotes the version of the model. Model updates are regularly released, and higher version numbers typically signify greater accuracy.

**dna\_r10.4.1\_e8.2\_400bps\_sup@v5.2.0**



# How to use stand-alone Dorado -> simplex basecalling

- First, download Dorado from [www.github.com/nanoporetech/dorado](https://www.github.com/nanoporetech/dorado)
- Basecalling simplex data from a run ( e.g LSK114) using high -accuracy basecalling (hac)
  - dorado basecaller hac /path/to/pod5/data > calls.bam

```
(base) steven.batinovic@X4NVGK4F41 ~ % dorado basecaller hac pod5/DNA/ > calls.bam
[2024-03-26 06:38:30.366] [info] Assuming cert location is /etc/ssl/cert.pem
[2024-03-26 06:38:30.367] [info] - downloading dna_r10.4.1_e8.2_400bps_hac@v4.3.0 with httplib
[2024-03-26 06:38:31.306] [info] > Creating basecall pipeline
[2024-03-26 06:38:37.235] [info] - set batch size to 480
[2024-03-26 06:39:55.330] [info] > Simplex reads basecalled: 4000
[2024-03-26 06:39:55.330] [info] > Basecalled @ Samples/s: 1.059702e+06
[2024-03-26 06:39:55.394] [info] > Finished
```



# How to use stand-alone Dorado -> methylation detection

- What if you would like to pull out base modification data during basecalling?
- Basecalling simplex data using high -accuracy basecalling (hac) with 5mC and 5hmC detection (in CpG contexts)
  - dorado basecaller hac,5mCG\_5hmCG /path/to/pod5/data > calls.bam

```
[base] steven.batinovic@X4NVGK4F41 ~ % dorado basecaller hac,5mCG_5hmCG pod5/DNA/ > calls.bam
[2024-03-26 06:58:45.227] [info] Assuming cert location is /etc/ssl/cert.pem
[2024-03-26 06:58:45.228] [info] - downloading dna_r10.4.1_e8.2_400bps_hac@v4.3.0 with httpLib
[2024-03-26 06:58:46.400] [info] - downloading dna_r10.4.1_e8.2_400bps_hac@v4.3.0_5mCG_5hmCG@v1 with httpLib
[2024-03-26 06:58:47.478] [info] > Creating basecall pipeline
[2024-03-26 06:58:52.977] [info] - set batch size to 480
[2024-03-26 07:00:32.501] [info] > Simplex reads basecalled: 4000
[2024-03-26 07:00:32.501] [info] > Basecalled @ Samples/s: 8.315003e+05
[2024-03-26 07:00:32.650] [info] > Finished
```

dorado basecaller sup@v5.2.0,m5C\_2OmeC,inosine\_m6A\_2OmeA,pseU\_2OmeU,2OmeG pod5s/ > calls.bam





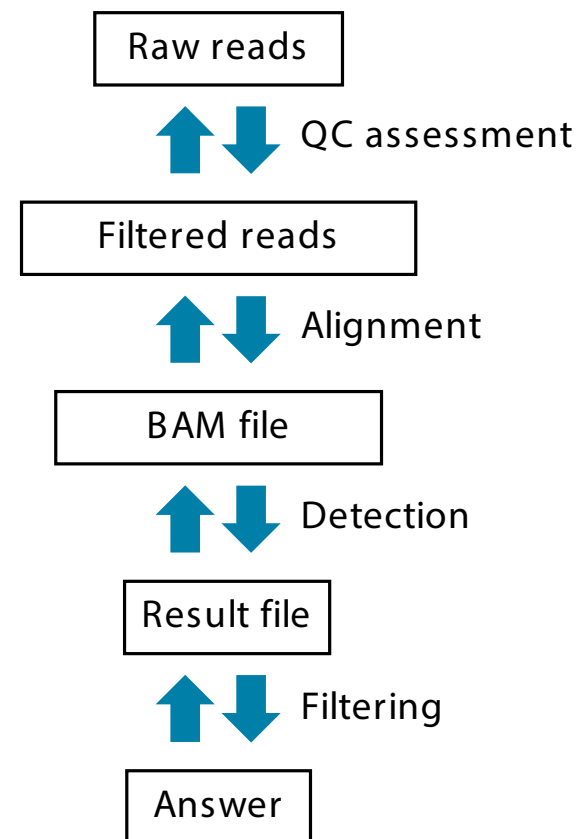
# Bioinformatics workflows

## Bioinformatics features many file formats

- Raw data (POD5, FASTQ, BAM)
- Alignment (SAM/BAM/)
- Reference genomes (FASTA)
- Annotation files (GTF/GFF/BED)
- Result files (VCF/TSV/BCF/)

## Automated bioinformatics

- Pipelines are multi-step analysis workflows
- May involve many software
- Analyses can be automated and run using single command
- Automatically use output from one command as input to another
- Generate reports and results in reproducible way



# Thank you

---

Oxford Nanopore Technologies, the Wheel icon, AmPORE-TB, ElysION, EPI2ME, Flongle, GridION, MinION, MinKNOW, PromethION, and TraxION are registered trademarks or the subject of trademark applications of Oxford Nanopore Technologies plc in various countries. Information contained herein may be protected by patents or patents pending of Oxford Nanopore Technologies plc. All other brands and names contained are the property of their respective owners.

© 2025 Oxford Nanopore Technologies plc. All rights reserved.

Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.