

PromethION 2 Integrated IT requirements

Overview

The PromethION 2 Integrated (P2i) is a benchtop device for nanopore sequencing designed to run and analyse up to two flow cells. It is ideal for labs with multiple projects that need the advantages of nanopore sequencing:

- Simple library preparation
- Real-time analysis
- Biological insights from long reads

In addition, the P2i enables users to offer nanopore sequencing as a service when certified.

The device benefits from the inclusion of on-board compute which permits device control, data acquisition, real-time basecalling and data streaming, all without placing any additional burden on existing IT infrastructure. It can also perform post-basecalling data analysis, e.g. using EPI2ME workflows.

All device control, data acquisition and basecalling on the device is carried out by pre-installed custom software created by Oxford Nanopore Technologies. The default data analysis workflow when using the P2i is as below:



Figure 1: Default data analysis workflow of the PromethION 2 Integrated device

Specifications

The P2i is designed around a simple user interface on top of cutting-edge custom electronics providing real-time analysis solutions. It has a built-in touchscreen that enables you to start and monitor sequencing runs. You can also use an external monitor to use the integrated computer for downstream analysis such as EPI2ME.

Component	Specification
Size and weight	180 mm x 225 mm x 430 mm, 10.6 kg
Environmental conditions	Designed to sequence at +18°C to +25°C (Functional range of electronics +5°C to +40°C)
Display output	1x HDMI 2.0 Port (up to 4K resolution at 60 Hz) 1x Display Ports
USB ports	4x USB 3.0 Type-A ports (up to 10 Gb/s)

Component	Specification
Networking	1x 2.5 Gb/s Ethernet (RJ45 connector)
Integrated touch-screen display	5.5" (diagonal) AMOLED touch screen display
Audio	1x 3.5 mm audio output (top)
Power	750 W power supply
Storage	15 TB SSD
Memory	64 GB DDR4
CPU/GPU	1x Intel Core i7 (12-core/20-threads) 1x NVIDIA Ampere-series GPU
Operating system	Ubuntu 20.04 LTS
Software installed	MinKNOW
Telemetry feedback	HTTPS/port 443 to 52.17.110.146, 52.31.111.95, 79.125.100.3 (outbound-only access) or DNS rule for ping.oxfordnanoportal.com
EPI2ME analysis	Ethernet: HTTPS/port: 443 TCP access to AWS eu-west-1 IP ranges: http://docs.aws.amazon.com/general/latest/gr/aws-ip-ranges.html
Software updates	HTTPS/port 443 to 178.79.175.200 and 96.126.99.215 (outbound-only access) or DNS rule for cdn.oxfordnanoportal.com

Telemetry

MinKNOW collects telemetry information during sequencing runs as per the Terms and Conditions to allow monitoring of device performance and enable remote troubleshooting. Some of this information comes from free-form text entry fields, therefore no personally-identifiable information should be included. We do not collect any sequence data.

The EPI2ME platform is hosted within AWS and provides cloud-based analysis solutions for multiple applications. Users upload sequence data in FASTQ format via the EPI2ME Agent, which processes the data through defined pipelines within the EPI2ME Portal. Downloads from EPI2ME are either in Data+Telemetry or Telemetry form. The EPI2ME portal uses telemetry information to populate reports.

Software updates

The IP address from which you receive software updates will depend on your geographical location. You can update through the software UI or through the advanced package tool (apt) that is used to update software on Linux-based machines. This is preinstalled on the P2i and available through the Terminal application. To update via apt, you require outbound-only access. We notify users about software updates through the Nanopore Community and provide full instructions for updating in each release note.

Storage

File types

The nanopore application software, MinKNOW, can output sequencing data in three file types: POD5, FASTQ and BAM. Basecalling summary information is stored in a sequencing_summary.txt file:

- FASTQ is a text-based sequence storage format, containing both the sequence of DNA/RNA and its quality scores. By default, nanopore sequencing experiments save up to 4000 DNA sequences in one FASTQ file. File size can vary from <1 Mb to tens of

Mb depending on the number and length of sequences.

- BAM files are output if you perform alignment or modified base calling on the basecalled dataset. BAM is an industry-standard file format for storing alignment and modified base information that is typically far smaller than raw .pod5 data. Full specification is available on the [Samtools GitHub page](#).
- POD5 is an Oxford Nanopore-developed file format which stores raw nanopore data in an accessible way and replaces the legacy .fast5 format. This output also reads and writes data faster, uses less compute and has smaller raw data file size than .fast5.
- .fast5 is a legacy file format based upon the .hdf5 file type, which contains all information needed for analysing nanopore sequencing data and tracking it back to its source. A .fast5 file contains data from multiple reads (4000 reads as default), and is several hundred Mb in size.
- sequencing_summary.txt contains metadata about all basecalled reads from an individual run. Information includes read ID, sequence length, per-read q-score, duration etc. The size of a sequence summary file will depend on the number of reads sequenced.

Example file sizes below are based on different throughputs from an individual flow cell, with a run saving POD5, FASTQ, and BAM files with a read N50 of 23 kb. TMO = theoretical maximum output.

Flow cell output (Gbases)	POD5 storage (Gbytes)	FASTQ.gz storage (Gbytes)	Unaligned BAM with modifications (Gbytes)
100	700	65	60
200	1,400	130	120
290 (TMO)	2,030	188.5	174

As an experiment progresses, POD5 files are produced for all reads as default. If you choose to basecall your data, the MinKNOW software uses POD5 files to generate sequence data which it then stores in FASTQ files and/or BAM files.

Data transfer and long-term storage

The P2i has sufficient SSD disk space for multiple sequencing experiments, storing POD5, FASTQ, and BAM data. However, it is imperative to clear this data store regularly to prevent successive runs from terminating due to lack of storage space. For this, a site must provide storage to transfer data off the device.

The P2i runs on Ubuntu and can mount multiple filesystem types. We recommend storage presented as NFS or CIFS. The form and volume of data to be stored will depend on your requirements:

- You can choose to store POD5 files with raw read data or delete them. If you wish to rebasecall your data at a future date, you can optionally save raw POD5 files as a toggle in MinKNOW.
- Retaining only FASTQ/BAM files will allow use of standard downstream analysis tools using the DNA/RNA sequence.

Change log

Date	Version	Changes made
24th April 2024	V3	Made some corrections to the values in "Specifications"
8th April 2024	V2	Corrected the environmental conditions to say "Designed to sequence at +18°C to +25°C"
10th October 2023	V1	Initial document introduction