PromethION Alpha-Beta

Site Installation and Device Operation Requirements

v0 Feb 2018

PromethION Alpha-Beta IT Requirements

Checklist

This checklist represents the minimal requirements for installing the PromethION in your institution. For full explanation of requirements please continue to the document. Please note either Fibre **OR** Copper connections can be utilized.

Item/Setup Required	Reason	Provided?
2x SFP+ modules supporting 10GBASE-SR or 10GBASE-LR **(not required if copper used)**	Connection to IT infrastructure and internet	
Appropriate 10 Gbps fibre cabling for chosen SFP+ modules **OR** Appropriate cabling for 10 Gbps copper	Connection to IT infrastructure and internet	
2x 10 Gbps fibre ports with DHCP service running **OR** 2x 10 Gbps copper ports with DHCP service running	Connection to IT infrastructure and internet	
HTTPS/port 443 TCP access to AWS eu-west-1 IP ranges listed at http://docs.aws.amazon.com/general/latest/gr/aws-ip-ranges.html	Telemetry feedback, EPI2ME analysis	
HTTPS/port 443 and HTTP/port 80 TCP access to Oxford Nanopore repository IP addresses 178.79.175.200, 96.126.99.215 and 106.187.40.102 for software distribution	Software updates	
Storage: *Sufficient infrastructure for required storage option	Streaming and long-term data storage	
USB mouse	Device control	
USB keyboard	Device control	
VGA compatible monitor and Power cable	Device control	
2x Power supply capable of delivering 2 kW and 500 W respectively	Device power	

*Size of storage required will depend on use case. Please refer to the information below for storage guidelines

PromethION Alpha-Beta Device

Overview

The PromethION Alpha-Beta is a benchtop device for nanopore sequencing designed to run and analyse up to 48 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 PromethION Alpha-Beta also allows users to offer nanopore sequencing as a service (following the end of the PEAP).

The PromethION Alpha-Beta benefits from the inclusion of on-board compute which permits data acquisition, analysis and feedback, basecalling, data streaming and device control, all without placing any additional burden on existing IT infrastructure.

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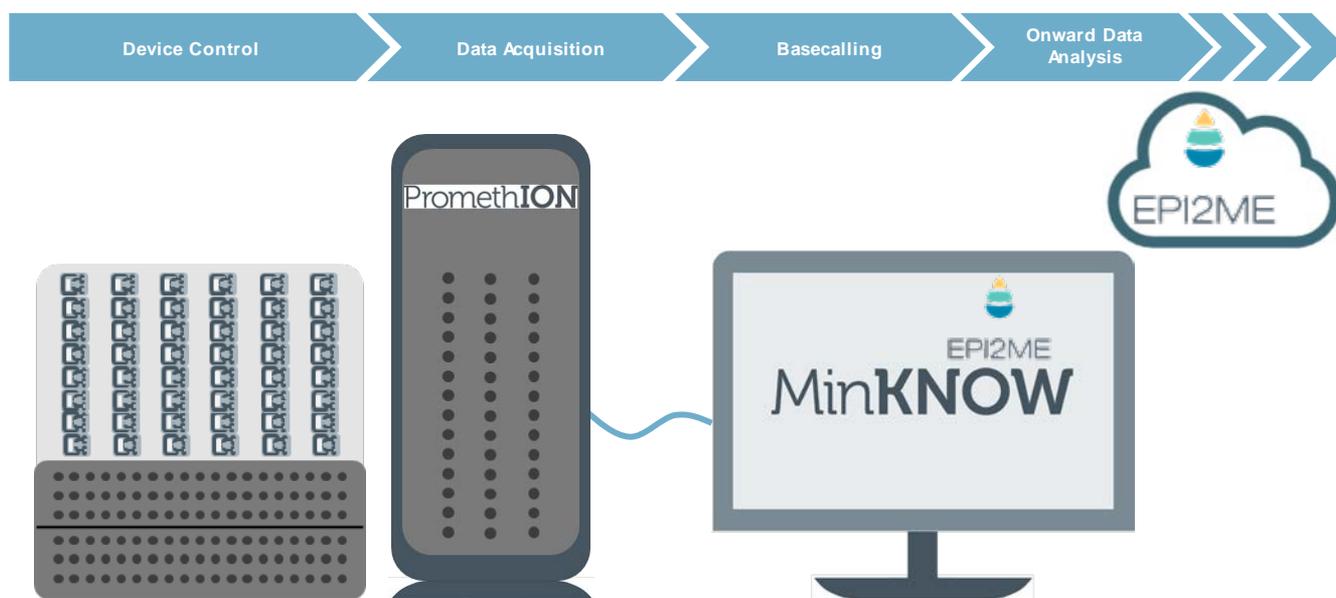

Figure 1: Default data analysis workflow of the PromethION Alpha-Beta device

Specifications

The PromethION Alpha-Beta is designed around a simple user interface on top of cutting-edge custom electronics providing real-time analysis solutions:

Component	Specification
Operating System	- Ubuntu 16.04 running off Intel CPU - Customers are advised to keep up to date with all software and security patches
Storage	- 28 TB internal SSD
Memory	- 384 GB RAM
Size and weight	Main Unit - **HxWxD** 220 x 437x 410 mm - **Weight** 40 kg Compute tower: - **HxWxD** 462 x 178 x 673 mm - **Weight** 28 kg
Environmental ranges	- -5° C to +40° C

Site Requirements

Installing the PromethION Alpha-Beta in your institution is similar to installing any new server. Requirements for the device are as follows:

Component	Requirements

2x 10 Gbps fibre port (SFP+ modules must be provided)
 	
OR
 	
2x 10 Gbps RJ45 Port
	- IP address via either DHCP service OR statically assigned
	- TCP running on port 80 HTTP and port 443 HTTPS
	- Requires Perimeter firewall permissions:
	  o To access AWS eu-west-1 IP ranges listed at https://ip-ranges.amazonaws.com/ip-ranges.json for telemetry feedback and EPI2ME analysis
	  o To 178.79.175.200, 96.126.99.215, and 106.187.40.102 for software updates
	- Connected to local storage infrastructure
	- **Optional:** For higher bandwidth connection into local infrastructure can bond either actively/passively, or using LACP

4x Power supply	- 2x for Compute Tower, C13 cables included with device
	  o Maximum power draw from Compute Unit of 2 kW
	- 1x for Sequencing Unit, C13 cable included with device
	  o Maximum power draw from Sequencing Unit of 500 W
	- 1x for monitor power supply

1x Monitor	- Compatible with VGA Port for device interaction

1x Keyboard	-	Compatible with USB connection for device interaction
1x Mouse	-	Compatible with USB connection for device interaction

Networking explanations

The two requirements for firewall permissions are:

1. Access to the AWS eu-west-1 region for telemetry feedback and use of the EPI2ME platform through port 443
2. Access to the Oxford Nanopore repository IP addresses 178.79.175.200, 96.126.99.215 and 106.187.40.102 for software distribution

Telemetry

Telemetry information is collected by MinKNOW sequencing runs as per the Terms and Conditions to allow monitoring of device performance and enable remote troubleshooting. Some of this information is obtained from free-form text entry fields, therefore no personally identifiable information should be included. No sequence data is collected.

The EPI2ME platform is hosted within AWS and provides cloud-based analysis solutions for multiple applications. Use of this platform entails users uploading sequence data in .fastq format from the EPI2ME Agent, which is then processed through defined pipelines within the EPI2ME Portal. Downloads from EPI2ME are either in Data+Telemetry or Telemetry form. Telemetry information is used to populate reports within the EPI2ME portal.

Software updates

Depending on your geographical region, only one of 178.79.175.200, 96.126.99.215 or 106.187.40.102 will be used for provision of updates to device software. Updates are triggered as pull requests, therefore outbound-only access is required.

Storage

File types

The PromethION Alpha-Beta stores biological data in two file types, .fast5 and fastq:

- .fast5 is a customized file format based upon the .hdf5 file type, which is designed to contain all information needed for analyzing nanopore sequencing data and tracking it back to its source. Each piece of DNA/RNA sequenced will have its own .fast5 file, with each positon on the device capable of producing millions of these in a single run. A.fast5 file is generally less than 1 Mb in size, however as length of DNA increases file size also increases

- .fastq is a universal text-based sequence storage format, containing both the sequence of DNA/RNA and its quality scores. By default the device saves up to 4000 DNA sequences in one .fastq file. File size can vary from < 1Mb to tens of Mb depending on number and length of sequences

As an experiment progresses .fast5 files are produced for all reads. If basecalling is chosen these reads are utilized by the onboard software (more information below) to generate sequence data which is then stored in .fastq files. While the .fast5 file format is capable of storing this sequencing data as well, this configuration is currently not utilized on the PromethION Alpha-Beta.

Long-term storage

The PromethION Alpha-Beta has sufficient SSD disk space for data to be written out in real time as both .fast5 and .fastq. However, heavy utilization of the device will rapidly lead to this on board storage being filled. Therefore it is imperative this data is streamed from the device in real-time in order to prevent parallel runs from terminating due to lack of storage space. For this a customer site must ensure connectivity to the local infrastructure is of sufficient bandwidth to prevent data backing up. The PromethION Alpha-Beta offers two 10 Gbps ports for this purpose, with the customer able to choose between Ethernet or Fibre solutions.

The PromethION Alpha-Beta runs on Ubuntu and is able to mount multiple filesystem types. We recommend storage presented as NFS or CIFS. **This storage streamed to in real-time must be SSD** for its high write speed compared to HDD. The form (and volume) of data to be stored will depend on customer requirements:

- Storing .fast5 files with raw read data in will permit re-basecalling of data when new improvements are released by Oxford Nanopore
- Retaining just .fastq files will allow use of downstream analysis tools, but no further sequence data can be generated when improvements in basecalling become available.

Oxford Nanopore is unable to provide exact recommendations for storage as these will be site-specific. The above guidelines and requirements should be taken into consideration.

Included Software

Oxford Nanopore Technologies build and provide numerous software types involved in acquisition, orchestration and analysis:

- ⚙ MinKNOW
- ⚙ Albacore
- ⚙ Guppy
- ⚙ Dogfish
- ⚙ EPI2ME

MinKNOW

MinKNOW carries out several core tasks:

- Data acquisition
- Real-time analysis and feedback
- Data streaming
- Device control, including run parameter selection
- Sample identification and tracking
- Ensuring chemistry is performing correctly

It utilizes an intuitive GUI and receives updates on a regular basis. This is the core software provided by Oxford Nanopore without which the sequencing devices cannot be run. Data from MinKNOW is packaged into individual read .fast5 files. These .fast5 files are then utilized by other down-stream software.
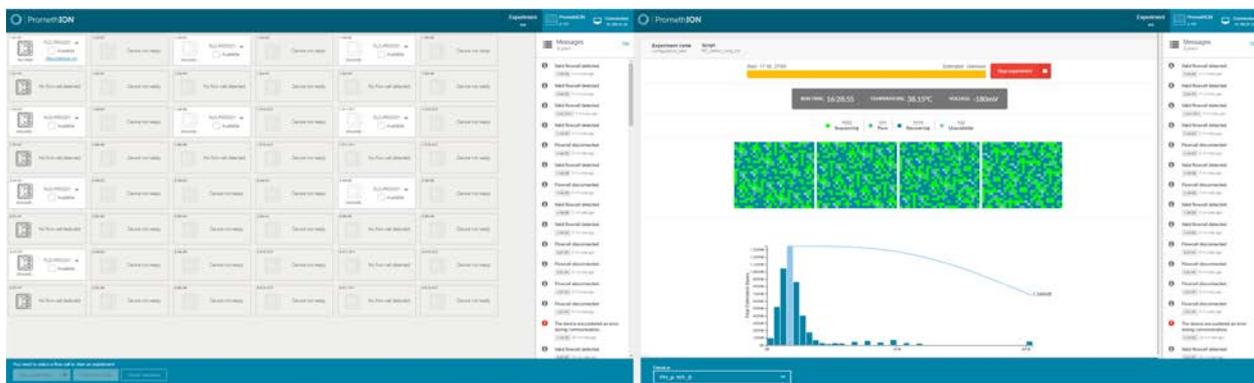

Figure 2: Example screen shots of the MinKNOW GUI. This permits selection and initiation of experiments, as well as providing real-time feedback on experiment progression

Albacore

Albacore is a production basecaller provided by Oxford Nanopore, and uses a command-line interface. It utilizes the latest in Recurrent Neural Network algorithms in order to interpret the signal data from the nanopore, and basecall the DNA or RNA passing through the pore. It implements stable features into Oxford Nanopore Technologies' software products, and is fully supported. It receives .fast5 files as an input, and is capable of producing:

- .fast5 files appended with basecalled information
- .fast5 files that have been processed, but basecall information present in a separate .fastq file

Guppy

Guppy is a production basecaller provided by Oxford Nanopore featuring the same core code as Albacore but optimized for running with basecall accelerators e.g. GPUs, FPGAs. Unlike Albacore, Guppy is not intended for direct interaction but instead makes use of Dogfish as a "broker" to divide processing of incoming .fast5 files across the available resource.

Dogfish

Dogfish is a service running on the PromethION Alpha-Beta that permits accelerated basecalling with Guppy, accessible through a command-line interface. It monitors for .fast5 files from MinKNOW then passes them into Guppy in a fashion that allows for utilization of the integrated basecalling accelerator.

EPI2ME

EPI2ME™ is an onwards data-analysis platform created by Oxford Nanopore's subsidiary company Metrichor. It provides users with real-time analysis such as species identification, alignment workflows and other bioinformatics solutions. It is currently provided as a cloud-based analysis platform which is initiated through the local EPI2ME Agent.
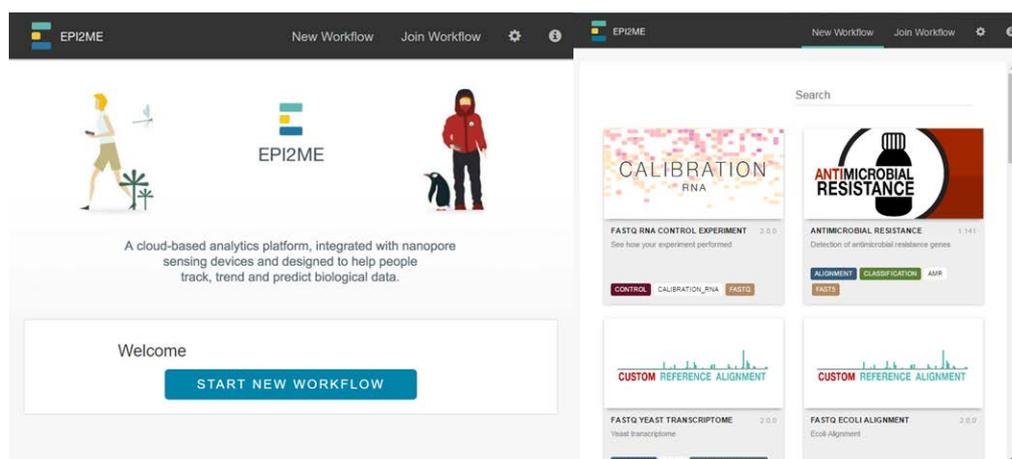

Figure 3: The EPI2ME Agent. Installed on the device, this allows selection and initiation of workflows within the EPI2ME portal.

Users will be able to review basic experimental statistics as well as use prepared workflows such as real-time taxonomic identification, reference alignment and many more.

Change log

Date	Version	Changes made
Feb 18	v0	Initial version