

MinION Computer Requirements

For Sequencing and Data Analysis



Oxford Nanopore Technologies
Oxford Science Park, Oxford OX4 4GA, UK
support@nanoporetech.com

www.nanoporetech.com

MinION Host Computer Requirements

Introduction

The MinION™ is a small self-contained, device for nanopore sequencing. It plugs directly into a standard USB 3.0 port on a desktop or laptop computer.

The MinION device is controlled by MinKNOW™ software.

MinKNOW carries out several core tasks: data acquisition; real-time analysis and feedback; basecalling; data streaming; device control including selecting the run parameters; sample identification and tracking and ensuring that the platform chemistry is performing correctly to run the samples.

Despite its size, the MinION is capable of sequencing many gigabases (Gb) in a single sequencing experiment. As such it is recommended that users purchase a powerful PC to ensure all the technology features are accessible.

MinION throughput example:

In Feb 2017 internal best >20 Gb was achieved internally in a 48 hour experiment

The default data-analysis workflow when starting to use the MinION is below:



EPI2ME™ is an onwards data-analysis platform, providing users with real-time analysis such as real-time species identification, alignment workflows and other bioinformatics solutions.

There are other onwards analysis pipelines available and used by the community. These are covered later on in the document.

Minimum host computer specification

This example will allow a user to run a MinION with real-time local basecalling. Given the amount of data potentially generated, the basecalling may fall behind and need to be completed at the end of an experiment.

Component	Configuration		
Operating Sytem	Windows	Mac	Linux
	7, 8, 10 (64 bit)	Yosemite	Ubuntu 14.4*
		El Capitan	
		Sierra (64 bit)	
Memory/RAM:	16GB RAM		
CPU	i7 or Xeon**		
Storage	1TB SSD		
Ports	USB3***		
<i>* Linux products are offered under limited support and may take the team longer to respond to queries</i>			
<i>** Users need to verify their i7 is a four core model or better.</i>			
<i>*** The MinION device is CE marked using USB3. If a user wished to use USBC they may but this invalidates the CE marking</i>			

Example computer configurations capable of keeping up with the MinION are provided in the **Data path** section of this document.

Host computer set up

The below requirements are to ensure correct operation of the MinION during a sequencing experiment:

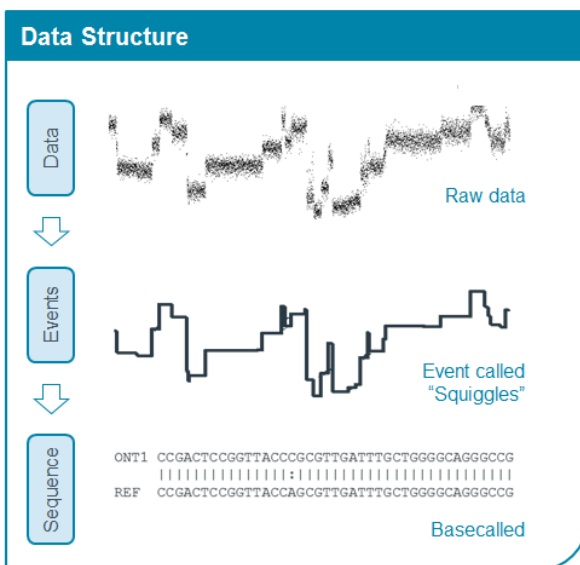
Component	Minimum Requirement
User Account Privilege Level	Local Administrator This is in order to give MinKNOW correct CPU priority for data acquisition.
Internet Connection	An internet connection is required at all times for software updates and telemetry Offline configurations can be made available for field use and expeditions. Please contact Support.
Firewall settings	Inbound traffic: MinKNOW downloads firmware from: mirror.oxfordnanoportal.com, resolving to americas.oxfordnanoportal.com, apac.oxfordnanoportal.com, europe.oxfordnanoportal.com depending on your geographical area. Outbound traffic: MinKNOW starts an experiment provided it can connect to the telemetry servers hosted on AWS. See https://ip-ranges.amazonaws.com/ip-ranges.json for your geographical area.
Antivirus settings	Antivirus software scanning the entire hard drive uses significant amounts of resources and may cause performance issues. For this reason it is recommended that the scans are switched to manual setting, rather than to automatic, so that they can be performed when the MinION is not in use.
OS update settings	OS system updates should be switched to manual mode as downloading them during a sequencing experiment may impact performance. Updates that require restart will fatally halt a run.
Web Proxy:	HTTP/S proxy support has been tested to the best of our ability.

MinION Basecalling

File Types

MinKNOW produces two types of .fast5 files:

- 1 - Raw data .fast5 files (typically less than 1MB in size).
- 2 - Basecalled .fast5 files (typically less than 1MB in size).



Event data is switched off in MinKNOW (1.5 onwards) by default. Users wishing to utilise event information to develop new analysis tools, or to use tools already developed, have two options for set up

- Configure MinKNOW to produce event data in the raw data .fast5 output file (these files are typically 8 times larger so ensure you have sufficient disk space at the start of a run)
- Re-run the raw data .fast5 through offline basecallers such as Nanonet or Albacore and utilise the output of these with event based tools.

Additionally from MinKNOW 1.6 onwards FastQ will be an optional file format.

Data Volume

Data volume example:

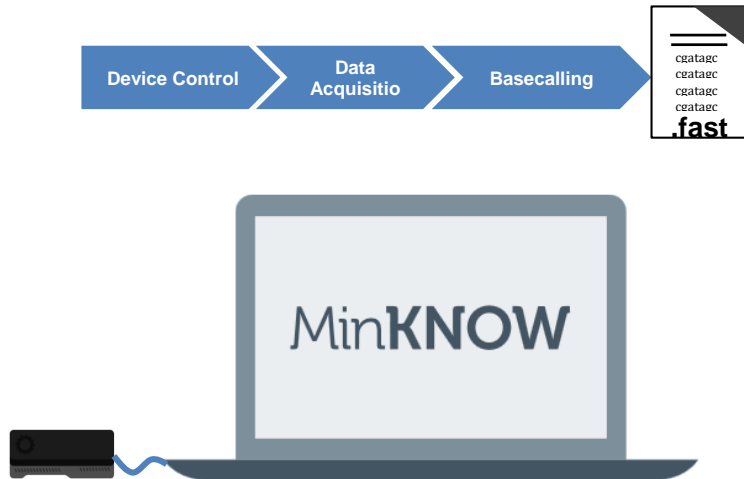
An Internal MinION run using MinKNOW1.5 produced >20Gb in 48 hours & generated ~ 250GB of .fast5 raw data. Additional storage (20 GB) was also in use for temporary files generated during the sequencing experiment and subsequent basecalling.

As many as 2 million read files have been produced during a sequencing experiment. These are stored in sub-folders containing 4000 reads each.

Large SSD disk storage is specified throughout documentation to cope with planned throughput increases and also to avoid experiments being halted by lack of disk space.

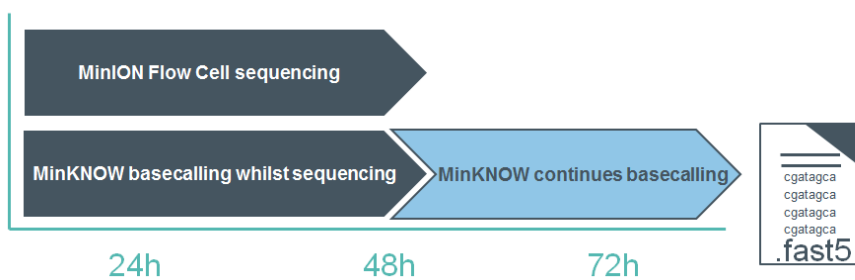
Data Path

MinKNOW Basecalling



When basecalling in MinKNOW read files will be organised into 3 final categories:

Pass	Reads have been produced and base called.
Skip	Reads have been generated and insufficient resources were available for basecalling. Reads are being held here for basecalling later on locally.
Fail	Reads have been produced and basecalled, but the basecalling was not completed successfully.



Users operating the MinION on a laptop will find that MinKNOW will basecall as much as possible in real time and then analyse the remaining files at the end of the sequencing experiment; the laptop will need to be left on for this to happen.

Example with MinKNOW 1.5:

A dual core i7 dual processor on a standard laptop will be able to basecall around 35,000 bases per second.

This equates to 2.5-3 Gb in 24 hours

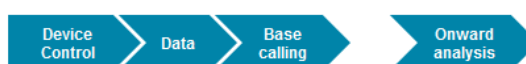
Higher specification machines can keep up with the MinION operating to its full throughput. Users considering applying nanopore to a production sequencing environment should consider GridION X5 or PromethION (please see nanoporetech.com for more information)

Local HPC for Basecalling

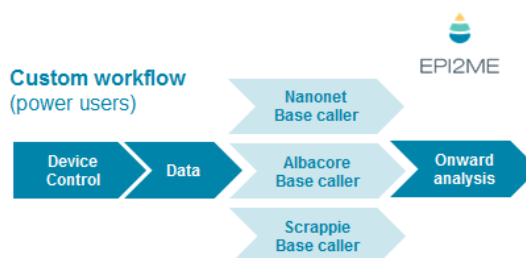
There are a number of local analyses options available to power users (more details in the community)

Name	Properties
Albacore	Production base caller <ul style="list-style-type: none"> - Implements stable features into product <i>Executables, MinKNOW, PromethION etc.</i> - Formal release and testing methods applied - Fully supported
Nanonet	Research base caller release <ul style="list-style-type: none"> - Available open source on GitHub - Was first release of RNN base caller - Limited support
Scrappie	Research base caller release <ul style="list-style-type: none"> - Available to Nanopore Developers on GitHub - Cutting edge Research base caller <ul style="list-style-type: none"> - Currently contains transducer algorithm aimed at reporting homopolymer regions - Limited Support

Standard workflow



Custom workflow (power users)



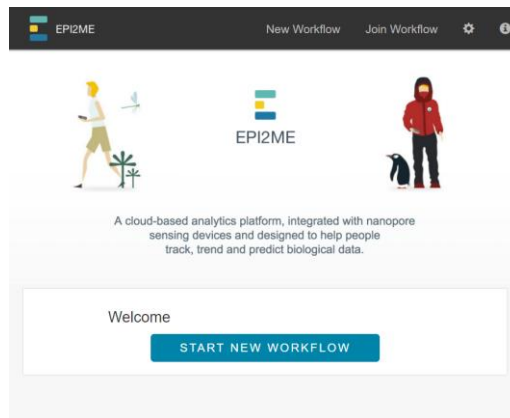
If a user wishes to supplement their host PC with desktop or server based basecalling solutions some example configurations are provided.

High Performance Desktops	Infrastructure (Servers)
DELL Precision Tower 7910 <ul style="list-style-type: none"> - 2 x Dual Intel Xeon Processor (10 cores each at 3Ghz +) - 64GB RAM (8x8GB) - 1TB SSD Hard Drive (can place more in the configuration) <p><i>High Performance Desktop can keep up with 8-10 Gb 1D basecalling /day</i></p>	DELL Power Edge R730 <ul style="list-style-type: none"> - 2 x Dual Intel Xeon Processor (18 cores or more) - 128GB RAM - 2 x 512GB SSD Hard Drive - Additional storage for run data storage (recommend 8TB all together) <p><i>With a machine specified to the above we expect 1D basecalling to be run at 20Gb / day</i></p> <p><i>Users still using 2D chemistry will find that with a machine specified to the above we expect 2D basecalling to be run at 4Gb / day. 2D Chemistry is due to be discontinued by Oxford Nanopore in May 2017.</i></p>

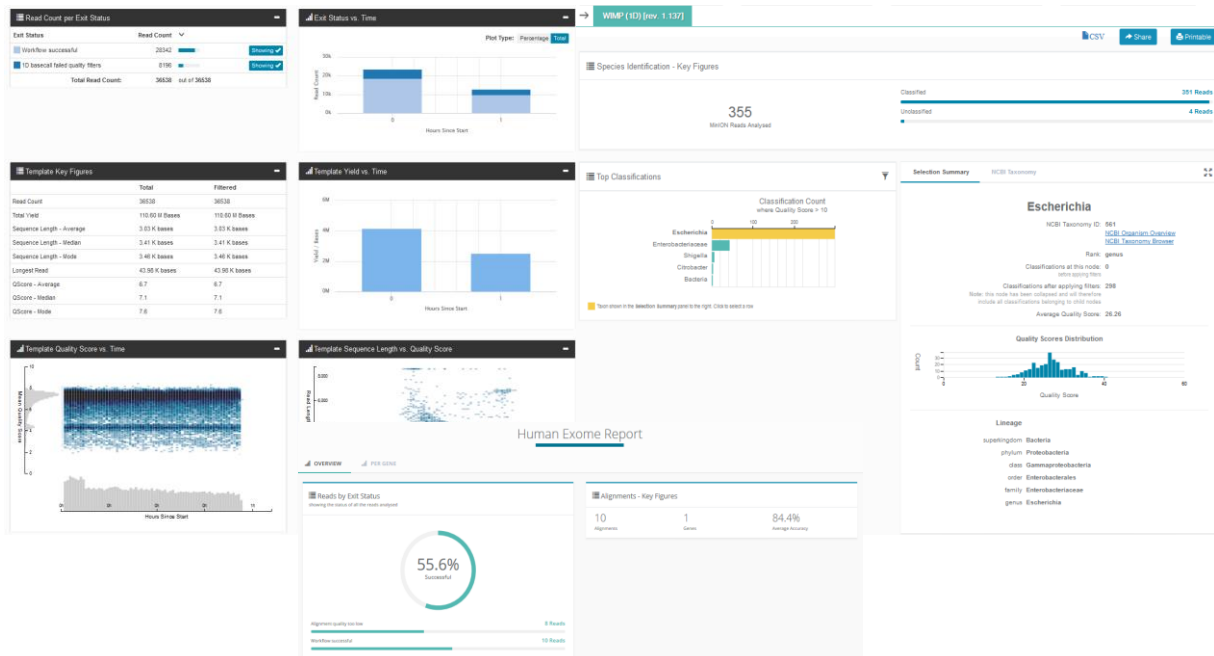
Onwards Analysis

Nanopore provided

Part of the getting started process for a new user is the creation of an EPI2ME account. EPI2ME is a cloud based analysis platform that generates, in real time, run metrics and carries out rapid bioinformatics workflows.



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.



Community tools

There are a large number of onwards analysis tools developed by the Nanopore Community. Many of these are publically visible in the publications section of nanoporetech.com or for community members...simply log in and start the conversation.

