Supplementary information

Towards population-scale long-read sequencing

Supplementary table 1. Comparison of three strategies for population-scale, long-read sequencing projects.

Approach	Feature	Number
General parameters	Cohort size (individuals or samples)	2,500
	Genome size	3 Gbp
	PromethION flow cell yield	100 Gb
	NovaSeq S4 PE150 lane yield	800 Gb
Strategy 1: Long-read sequencing of all samples at equal coverage ('full coverage' approach)		
Long-read sequencing	Number of samples	2,500
	Coverage	20-fold
	Data	150,000 Gb
	Number of PromethION flow cells	1,500 flow cells
Strategy 2: Long-read sequencing of a small proportion of samples at high coverage and the remaining samples at low coverage ('mixed coverage' approach)		
Long-read sequencing at high coverage	Number of samples	200 samples
	Coverage	30-fold
	Data	18,000 Gb
	Number of PromethION flow cells	180 flow cells
Long-read sequencing at low coverage	Number of samples	2,300 samples
	Coverage	8-fold
	Data	55,200 Gb
	Number of PromethION flow cells	552 flow cells
Total	Data needed	73,200 Gb
	Number of PromethION flow cells	732 flow cells
Strategy 3: Long-read sequencing of a small proportion of samples at high coverage and the remaining samples at high coverage with short-read sequencing ('mixed sequencing' approach)		

Long-read sequencing at high coverage	Number of samples	200 samples
	Coverage needed	30-fold
	Data needed	18,000 Gb
	Number of PromethION FC	180 flow cells
Short-read sequencing at high coverage	Number of samples	2300 samples
	Coverage needed	30-fold
	Data needed	207,000 Gb
	Number of NovaSeq6000 S4 lanes PE150	258.75 lanes