**\section{The Genomics Revolution}**

**\subsection{Illumina sequencing}**

%% Ion-torrent

%% 45

Pyro sequencing

%% sequencing cost, moore’s law

%% economics of scale

%% innovation on top of next-generation sequencing

%% ancient genomics

%% iteration of illumine sequencing machines: GAII

%% hi-c sequencing: loops configurations, topologically associating domain (TADs), A/B compartments

%% chip sequencing

%% 3C sequencing (Job Dekker)

The technical limitations of Illumina sequencing (base accuracy and short read length), however, has been the bottleneck for improving rare genetic disease diagnostics yield, detecting rare somatic mutations and constructing high-quality reference genomes for non-human species. De novo assembly of other species, previously, have been attempted using de brugjin graph based de novo assembly algorithms with short reads, but assemblies produced from short reads were highly fragmented and incomplete. In addition, scaffolding strategies often did not provide sufficient long-range information to produce chromosome-level pseudomolecules and as a result, these assemblies provided incomplete information for comparative genomics purposes. Hence, assemblies produced from short reads often have collapsed repeats or contigs that cannot be placed accurately. To construct complete assemblies, reads need to be longer than the repeats of the target genome such that the reads can traverse the repetitive regions and optimally have unique sequences flanking the repetitive sequences such that the read can be placed in the assembly graph unambiguously. Not all repetitive sequences are repetitive. There are unique class of repeats called segmental duplications, which doesn't have a classical repetitive sequence, has a unique sequence, but is duplicated across the many parts of the genome and are thought to be important in driving evolution and these segmental duplications are typically defined as sections greater than 1kb with sequence similarity above 90\% to other regions of the genome. To distinguish segmental duplications from one another, reads also need to have high base accuracy to be able to distinguish closest segmental duplications from one another. Long-reads from third-generation sequencing technologies such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) provide an alternative towards improving the rare genetic diagnostics yield and improving the reference genome qualities in terms of both completeness and contiguity. Long-reads produced from third-generation sequencing platforms were orders-of-magnitude longer than that from the Illumina platform but had a much higher error rate; 10-15\% error rate for continuous long reads (CLR) from PacBio and 20-35\% error rate for ONT reads. Because of these high error rates, higher sequencing costs (lower yield per dollar) and insufficient improvement in read length, these platforms had limited use except for rare cases for real-time monitoring of ... and de novo assembly of plants and animal genomes..., and detection of pathogenic mutations that could not be detected with short reads [ref, ref]. Despite high error rate, the longer read length enabled the detection of structural variations that could not be previously detected with short reads, doubled the number of structural variations that can be detected from a typical human genome compared to the human reference genome. The longer read length allowed for the de novo assembly of BAC clones to hierarchically assemble missing sequences, also known as gaps, in the human reference genome, which have been problematic to assemble before and reveal human-specific gene duplications.

%% population-genetics

%% population genomics

%% increase in the number of whole-genome sequenced with illumine sequencing

%% cancer genomics, driver mutation, mutational signatures

%% clinical sequencing

%% tumour evolution

%% liquid biopsy

The human genome project is estimated to have cost 3 billion dollars, equivalent to 1 dollar per base pair. And as technology becomes more ubiquitous and democratised, we have constantly shifted/moved from studying one individual to studying the group. We initially focused on studying a single individual and as sequencing cost has decreased, population genomic studies and the history of differences and going back in time to study our lineage. What is common and different.

Thanks to new technologies.

A wave of standardisation to create file formats that is universally accepted across the community.