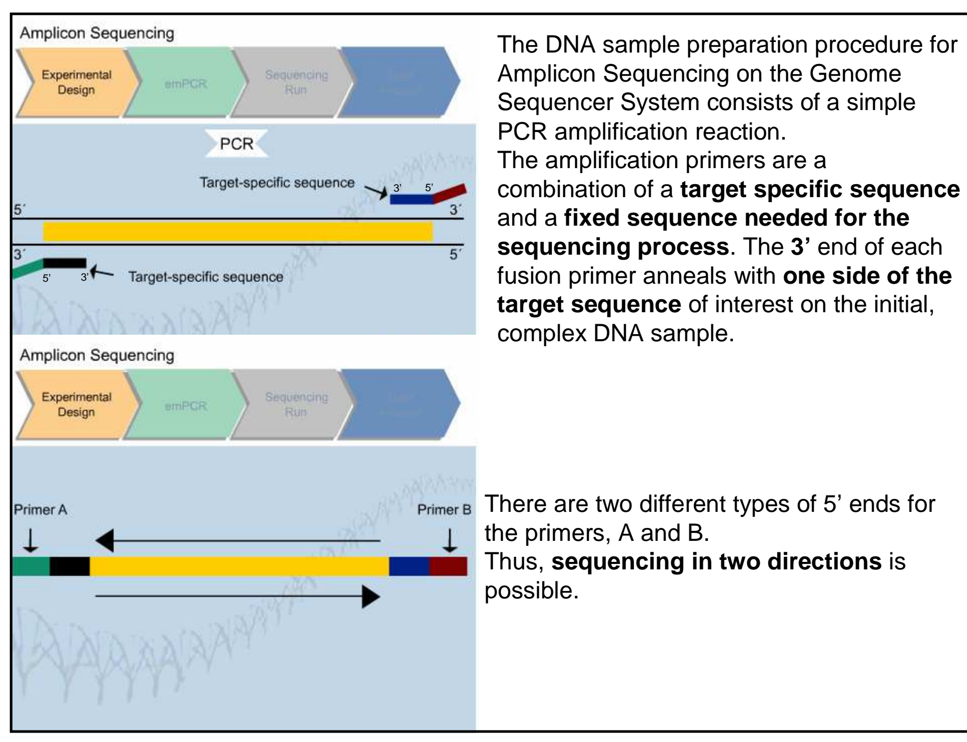
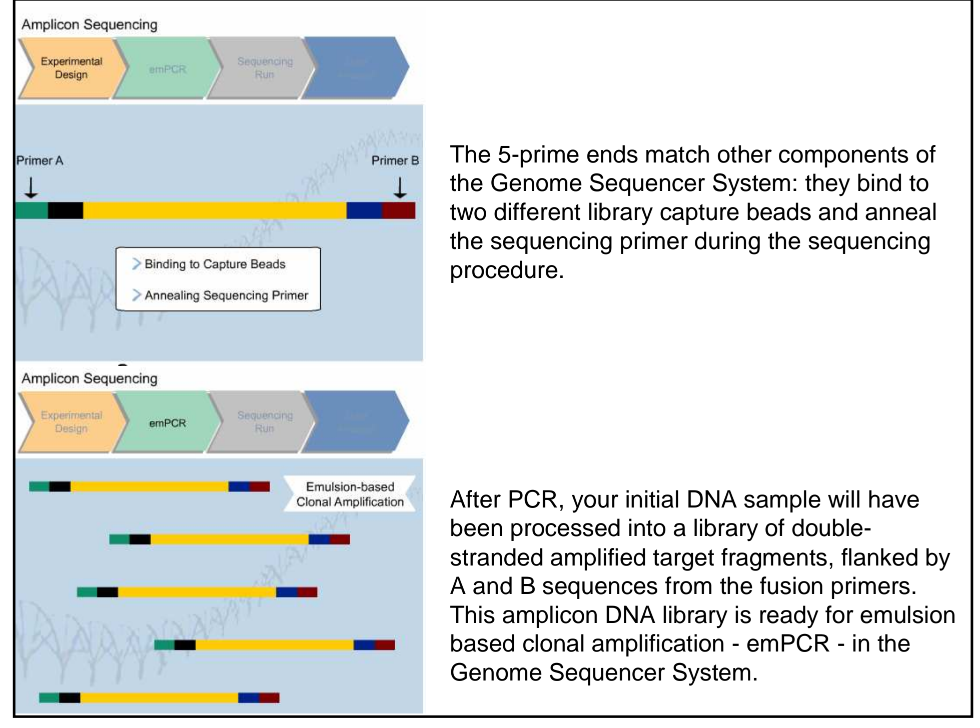


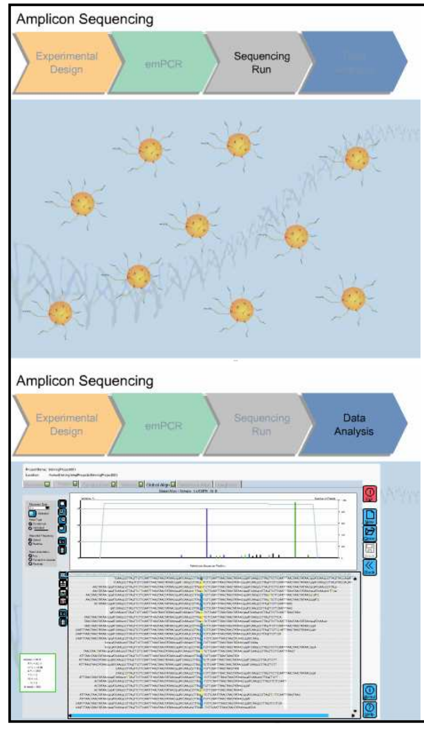
The DNA sample preparation procedure for Amplicon Sequencing on the Genome Sequencer System consists of a simple PCR amplification reaction.
The amplification primers are a combination of a **target specific sequence** and a **fixed sequence needed for the sequencing process**. The **3’** end of each fusion primer anneals with **one side of the target sequence** of interest on the initial, complex DNA sample.

There are two different types of 5’ ends for the primers, A and B.
Thus, **sequencing in two directions** is possible.



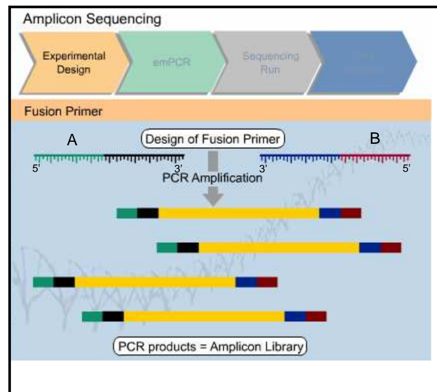
The 5-prime ends match other components of the Genome Sequencer System: they bind to two different library capture beads and anneal the sequencing primer during the sequencing procedure.

After PCR, your initial DNA sample will have been processed into a library of double-stranded amplified target fragments, flanked by A and B sequences from the fusion primers. This amplicon DNA library is ready for emulsion based clonal amplification - emPCR - in the Genome Sequencer System.

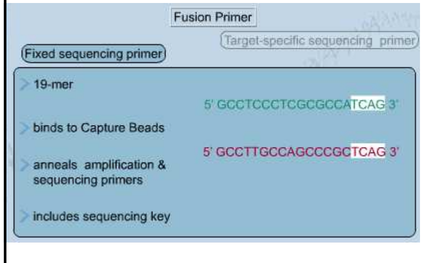


emPCR allows for single molecule sequencing without cloning the molecules of the PCR sample sequences into bacteria. The emPCR and sequencing reactions are performed following minor variations on the standard Genome Sequencer System procedures.

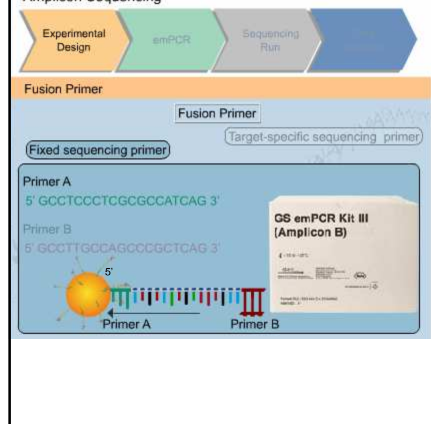
For data analysis, a complete software solution is available. The software provided is specifically tailored for the determination and quantification of sequence deviations. It makes available very powerful tools and a convenient interface able to identify rare variants, quantify their prevalence, and even verify whether variations observed at different sequencing positions are linked to each other, by looking at individual reads.



During the experimental design phase, **fusion primers** are designed for the generation of PCR products. **These PCR products, resulting from a simple PCR amplification step**, represent the Amplicon DNA library that contains the genomic spots that interest you. At the end of this experimental step, you will have an **Amplicon DNA library ready** for the amplification phase of the Genome Sequencer System.

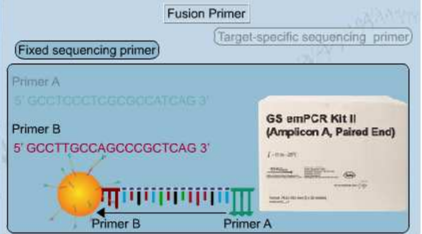


The primers used to generate amplicon libraries are composed of two sequence parts: a **fixed sequencing primer** and a **target-specific sequencing primer**.
The fixed sequencing primer at the 5' end is a Genome Sequencer System-specific **19-mer**. This primer **binds to the Capture Beads**, **anneals the emPCR amplification primers** and the sequencing primer, and ends with the **sequencing key "TCAG"**.

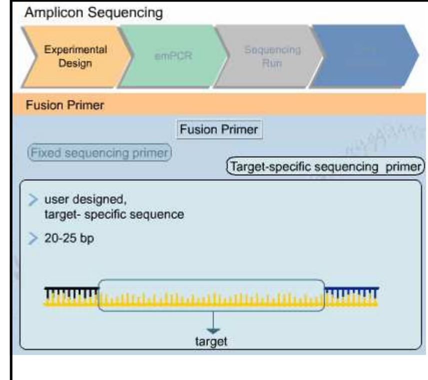


There are two types of these primers, termed Primer A and Primer B.

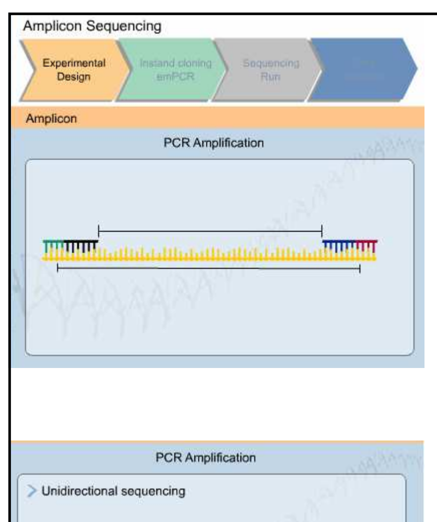
Primer A is complementary to the capture oligonucleotide which is covalently linked to the DNA Capture Beads in the GS emPCR Kit III. This kit allows you to sequence the amplified fragment from Primer B.



Primer B is complementary to the capture oligonucleotide which is covalently linked to the DNA Capture Beads in the GS emPCR Kit II. This kit allows you to sequence the amplified fragment from Primer A.

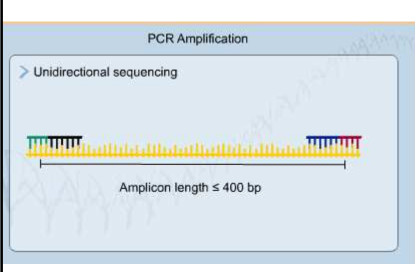


The **target-specific sequencing primer at the 3'end** is typically **20-25 base pairs** in length and designed to anneal to **either side** of the target to be sequenced. This requires some **knowledge of the target sequence**, and works best on target sequences that are not too variable.

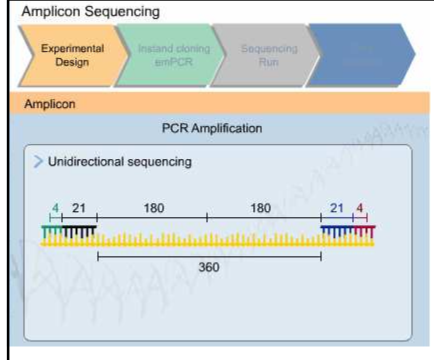


The current GS technology generates reads at an average length of greater than 200 base pairs. The distance between the two sequence-specific primers must be carefully considered if the "inserts" are to be sequenced in their entirety.

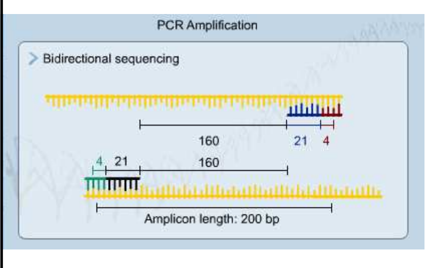
In general **we recommend bidirectional sequencing** to achieve high accuracy.



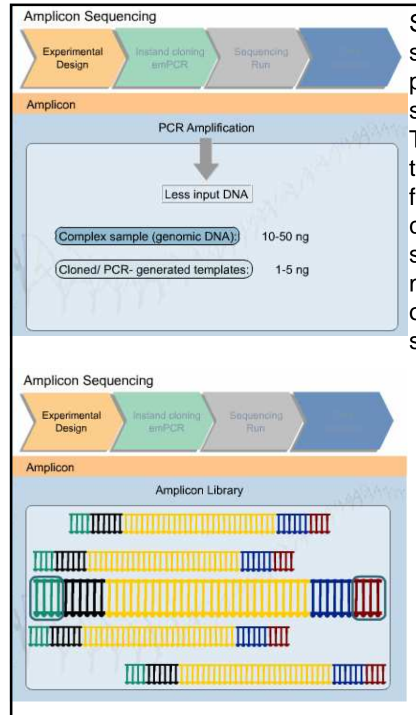
If unidirectional sequencing without any overlap is sufficient, the amplicon should be no longer than about 400 base pairs from key to key



For example, if the target-specific components of the Fusion Primers are each 21-mers, you will read about 180 nucleotides into the insert from each side - thus covering a 360 basepair insert.

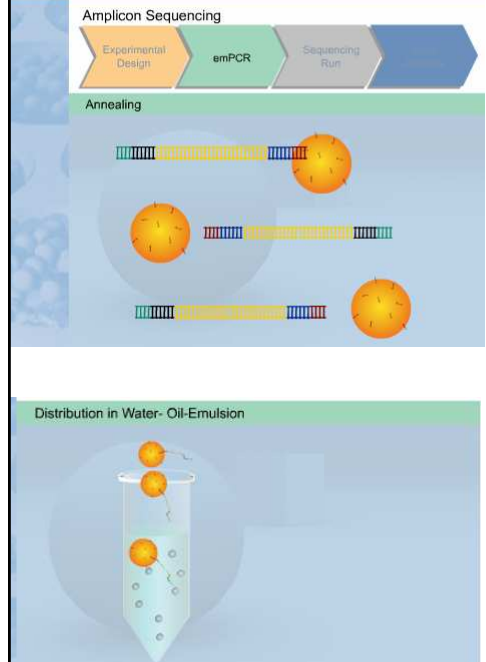


If bidirectional sequencing is required, the amplicon should be **no longer than about 200 basepairs from key to key**. With the same example of **21 basepair target-specific components of the Fusion Primers**, you can read **a 160 basepair insert from both ends**, so the amplicon can be this length plus the **target-specific primer** and the **key** on either side.



Since the preparation of a library for amplicon sequencing includes an amplification step, this procedure requires less input DNA than the standard sstDNA Library Preparation procedure. The amount required will depend on the nature of the experiment. For example, if you are searching for low abundance sequencing variants out of a complex sample, such as genomic DNA, you should start with 10 - 50 ng of DNA. If your starting material is cloned or **PCR-generated** template DNA of low to moderate complexity, then **1 - 5 ng** will suffice.

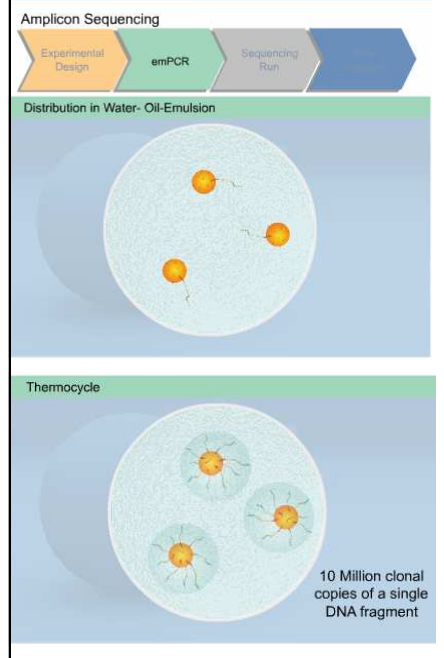
After successful completion of the experimental design phase, you will have an Amplicon DNA library ready for the amplification phase of the Genome Sequencer System. Your initial DNA sample has been processed into a library of **double-stranded amplified** fragments of a length determined by the **target-specific primer sequences you used**. The fragments are flanked with a **19 basepair Primer A** at one end and a **19 basepair Primer B** at the other.



For DNA amplification, the Genome Sequencer System employs emulsion based Clonal Amplification, emPCR, and thus avoids tedious and expensive cloning in *E.coli*.

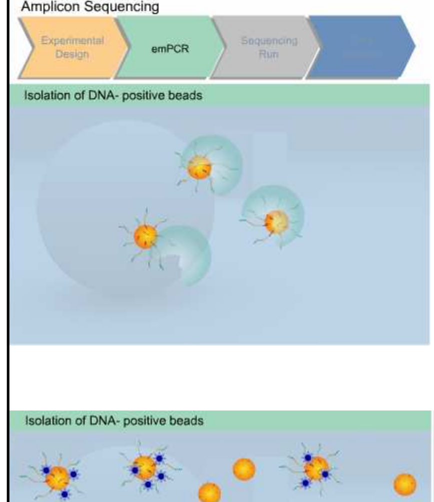
After purification and quantitation of the fragments, the amplicon DNA library is immobilized by annealing the amplicon DNA to an excess of DNA capture beads. Each bead carries no more than one amplifiable amplicon DNA molecule.

The entire bead-bound library is then emulsified with the amplification reagents in a water-in-oil mixture.



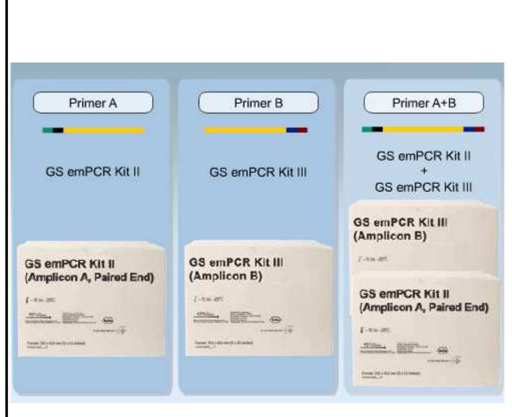
Each bead is captured within its own microreactor which contains the complete amplification mix and serves as an individual reaction micro-vessel for the amplification of a single amplicon DNA molecule.

All microreactors are processed in parallel. The amplification process results in beads covered with about **10 million clonal copies** of a single DNA fragment, each bead holding a different fragment.



After termination of the amplification steps, the microreactors are broken and DNA positive beads are enriched.

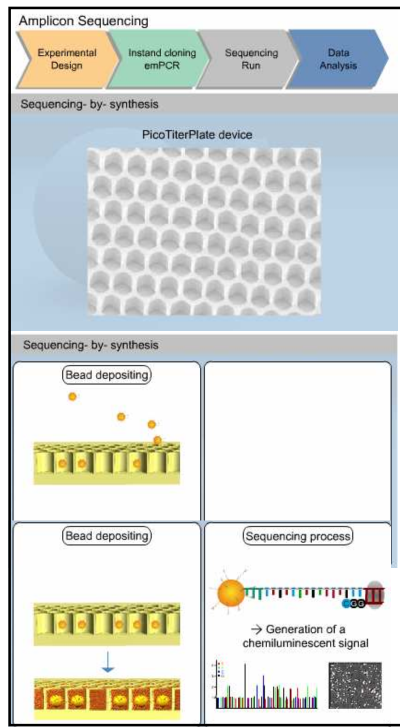
The purification is achieved by additing of magnetic enrichment beads that bind only to DNA positive beads.



Amplicon libraries can be read from either end, or from both the "A" and "B" primer ends.

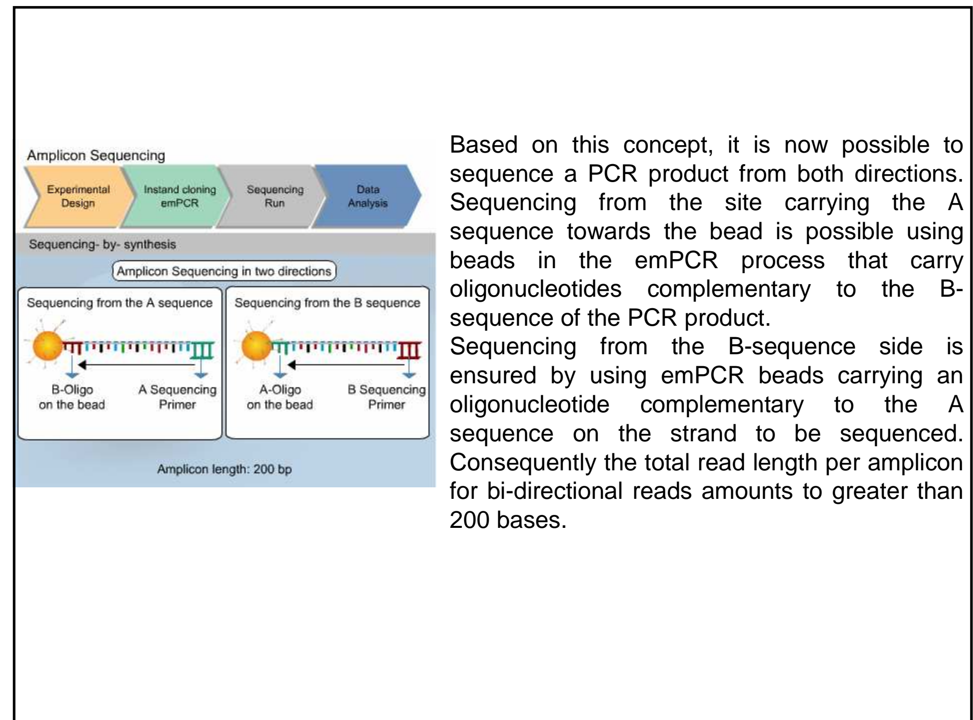
To read from Primer A, process the Amplicon library using the GS emPCR Kit II; to read from Primer B, process the library using the GS emPCR Kit III.

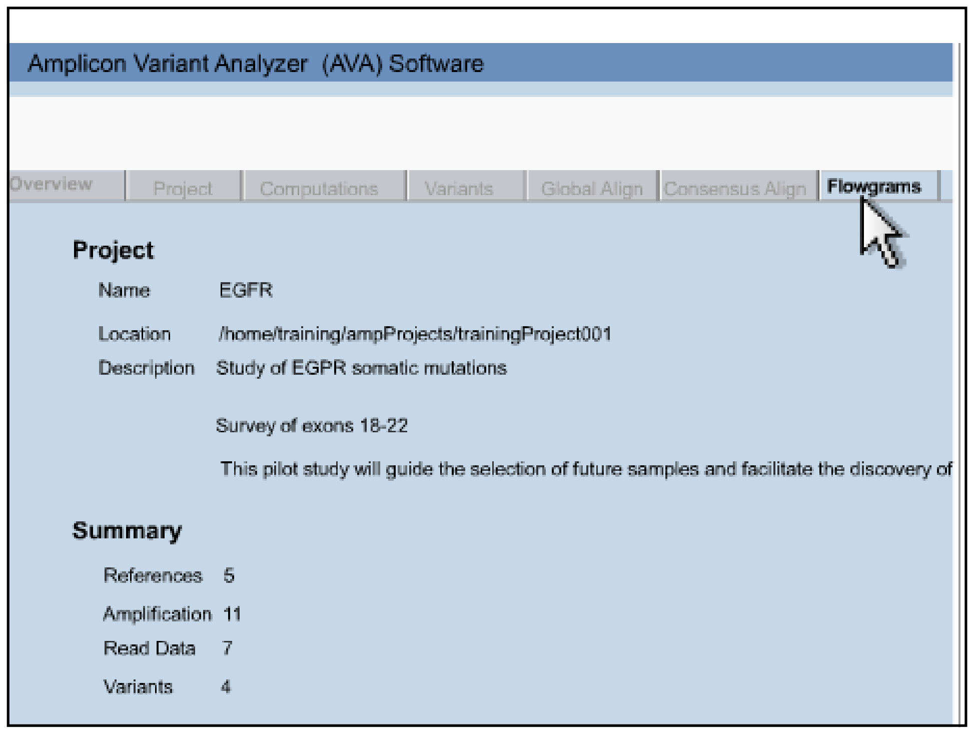
To read from both ends, **process an aliquot of the library using each of the GS emPCR Kits II and III** and **pool the two clonally amplified libraries before loading them onto the PicoTiterPlate device** for sequencing.

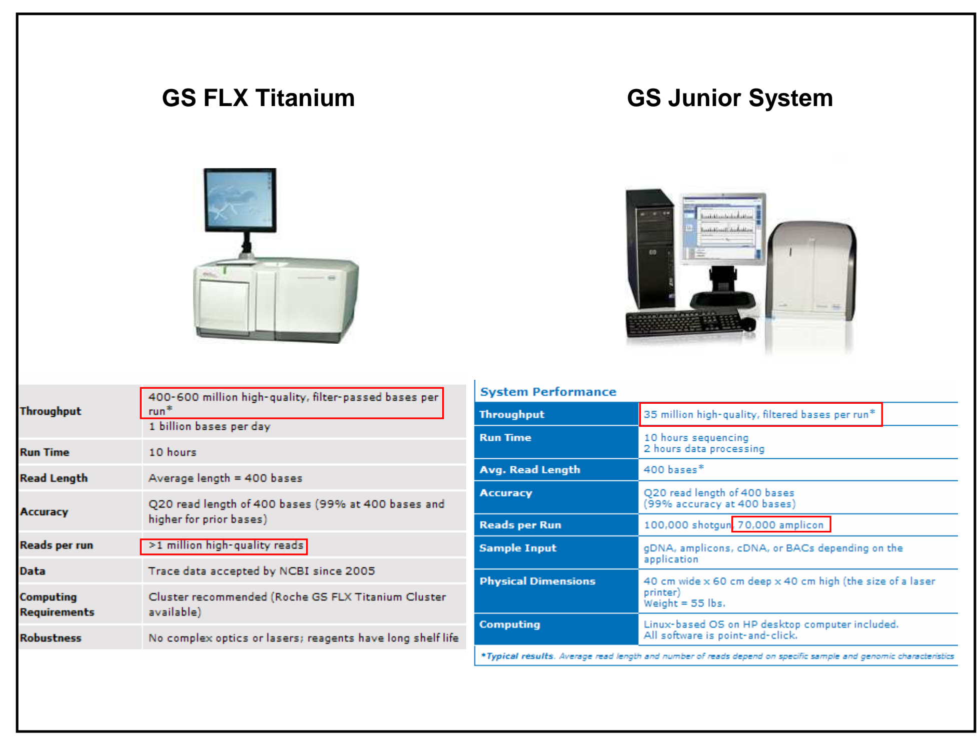


Another special feature of the Genome Sequencer System is its sequencing technology. Sequencing-by-synthesis is performed **simultaneously** in the **open wells** of a fiber-optic slide called a **PicoTiterPlate device**. For Amplicon Sequencing, and for all the applications of the Genome Sequencer System, sequencing-by-synthesis follows the same procedures.

First, the **DNA beads are loaded** together with **enzyme** and **packing beads** into the wells of the PicoTiterPlate device.
When the sequencing process starts, nucleotides are delivered to the wells in a **cyclically** repeating flow sequence. The sequencing-by-synthesis method is based on the generation of a chemiluminescent signal. The signal intensity at each nucleotide flow, for a particular well, indicates the number of nucleotides - if any - that were incorporated. If no nucleotide is incorporated during a nucleotide flow, no light signal is generated, and the complementary strand is not elongated.


Amplicon Sequencing
Experimental Design
Instand cloning emPCR
Sequencing Run
Data Analysis
Sequencing- by- synthesis
Amplicon Sequencing in two directions
Sequencing from the A sequence
Sequencing from the B sequence
B-Oligo on the bead
A Sequencing Primer
A-Oligo on the bead
B Sequencing Primer
Amplicon length: 200 bp
Based on this concept, it is now possible to sequence a PCR product from both directions. Sequencing from the site carrying the A sequence towards the bead is possible using beads in the emPCR process that carry oligonucleotides complementary to the B-sequence of the PCR product.
Sequencing from the B-sequence side is ensured by using emPCR beads carrying an oligonucleotide complementary to the A sequence on the strand to be sequenced. Consequently the total read length per amplicon for bi-directional reads amounts to greater than 200 bases.


Amplicon Variant Analyzer (AVA) Software
Overview
Project
Computations
Variants
Global Align
Consensus Align
Flowgrams
Project
Name EGFR
Location /home/training/ampProjects/trainingProject001
Description Study of EGPR somatic mutations
Survey of exons 18-22
This pilot study will guide the selection of future samples and facilitate the discovery of
Summary
References 5
Amplification 11
Read Data 7
Variants 4



Throughput	400-600 million high-quality, filter-passed bases per run* 1 billion bases per day
Run Time	10 hours
Read Length	Average length = 400 bases
Accuracy	Q20 read length of 400 bases (99% at 400 bases and higher for prior bases)
Reads per run	>1 million high-quality reads
Data	Trace data accepted by NCBI since 2005
Computing Requirements	Cluster recommended (Roche GS FLX Titanium Cluster available)
Robustness	No complex optics or lasers; reagents have long shelf life

System Performance

Throughput	35 million high-quality, filtered bases per run*
Run Time	10 hours sequencing 2 hours data processing
Avg. Read Length	400 bases*
Accuracy	Q20 read length of 400 bases (99% accuracy at 400 bases)
Reads per Run	100,000 shotgun 70,000 amplicon
Sample Input	gDNA, amplicons, cDNA, or BACs depending on the application
Physical Dimensions	40 cm wide x 60 cm deep x 40 cm high (the size of a laser printer) Weight = 55 lbs.
Computing	Linux-based OS on HP desktop computer included. All software is point-and-click.

****Typical results**. Average read length and number of reads depend on specific sample and genomic characteristics*