

# Next-Gen Bioinformatics



## Analyses & visualisation of sequence data from Roche GS FLX, Illumina Genome Analyzer & ABI 3730XL

Dr. Kerstin A. Stangier<sup>1</sup>, Dipl.-Bioinf. Ulrike Schöck<sup>1</sup>, Verena Tischler<sup>2</sup>, Dr. Yadhu Kumar<sup>1</sup>, Dr. Christopher Bauser<sup>1</sup>

<sup>1</sup> GATC Biotech AG, Jakob-Stadler-Platz 7, 78467 Konstanz, Germany, customerservice@gatc-biotech.com

<sup>2</sup> Fachhochschule Gelsenkirchen, Fb Angewandte Naturwissenschaften, August-Schmidt-Ring 10, 45665 Recklinghausen, Germany

### Ultra High-Throughput Sequencing

The technical capability of the Next-Gen sequencing systems, particularly embodied by the Roche GS FLX and the Illumina Genome Analyzer, generate huge amounts of sequence information:

#### Roche GS FLX

Read length Ø 100 bp and Ø 250 bp  
Mate pairs 2x up to 100 bp insert sizes up to 2 kb  
No. of reads/run up to 400,000

#### Illumina Genome Analyzer (GA)

17 bp, 25 bp and 36 bp  
2 x 25 bp or 2 x 36 bp insert sizes up to 600 bp  
up to 60,000,000

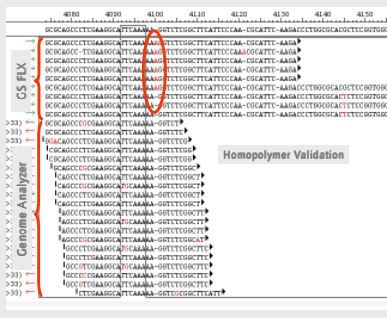
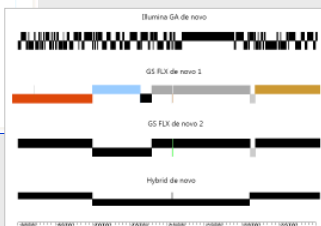
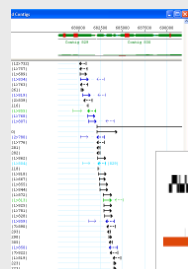
#### ABI 3730XL

up to 1,100 bases  
96 reads / run sizes  
approx. 1 MB per day

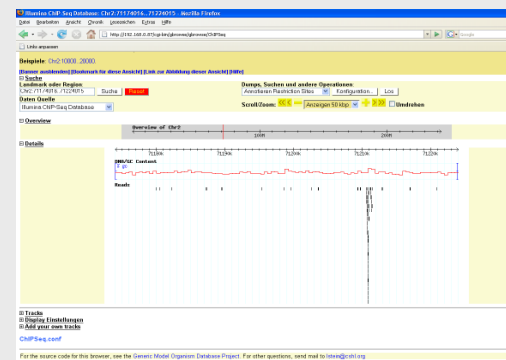
### Proprietary Data Analysis

GATC offers a wide range of bioinformatic solutions for genome assembly, transcriptome analysis or small RNA analysis. Proprietary tools are used for the analysis and handling of next-gen sequencing data, while third party tools are available for optimised *de novo* assembly of sequence data from the Genome Analyzer (GA) or improved hybrid assemblies.

#### De Novo / Hybrid Assembly

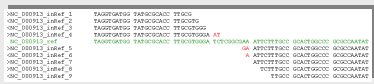


#### Transcriptome/ChIP Seq Analysis

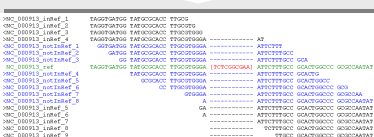


### Resequencing

#### SNP/Deletion Discovery



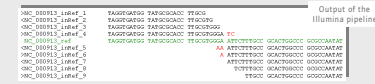
Subsequent analysis with proprietary GATC tool



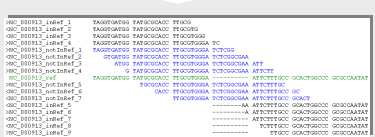
Green: Reads have expected insert sizes

#### Paired End Analysis

#### SNP/Insertion Discovery



Subsequent analysis with proprietary GATC tool



Red: Deviation from expected insert sizes

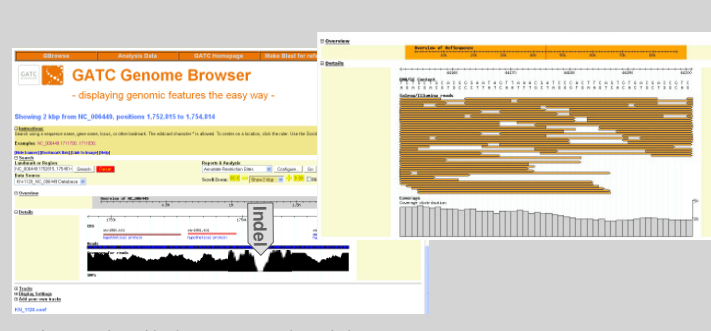
### Data Assembly and Visualisation

SeqMan Genome Assembler™ (SMGA; DNASTAR) assembles Illumina, Roche & Sanger data. Data files from SMGA can be imported into SeqMan Pro™ to permit a wide range of sequence analyses and visualisations. SeqMan Pro™ displays electropherograms as well as flowgrams.



### Data Visualisation in GATC Genome Browser

The GATC Genome Browser\* desktop application allows a convenient overview of e.g. whole genome *de novo* or re-sequencing, ChIP or SAGE experiments. The visualisation of the coverage helps to identify InDels and rearrangements within the genome.



\* Based on GBrowse; Stein LD et al. (2002). Genome Res 12: 1599-610 (www.genmud.org)