

Dr. Kerstin A. Stangier | Dipl.-Bioinf. Ulrike Schöck | Dr. Yadhu Kumar | Dr. Christopher Bauser  
 GATC Biotech AG, Jakob-Stadler-Platz 7 | 78467 Konstanz, Germany | customerservice@gatc-biotech.com

### 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 II, generate huge amounts of sequence information:

	Roche GS FLX (incl. Titanium)	Illumina Genome Analyzer II (GA)
Read length	Ø 100 bp, Ø 250 bp & Ø 400 bp	36 bp, 50 bp & 75 bp
Mate pair libraries	insert sizes up to 10 kb	insert sizes up to 10 kb
No. of reads/run	> 1,000,000	up to 80,000,000

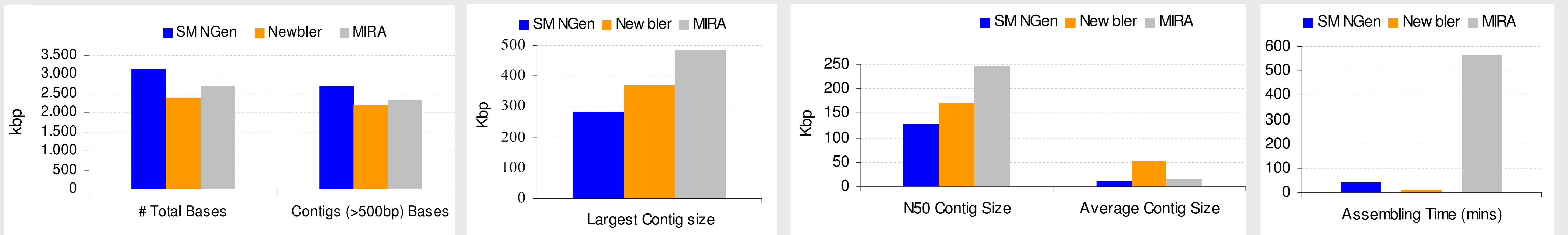


### Proprietary Data Analysis

GATC offers a wide range of bioinformatic solutions for genome assembly and alignment, transcriptome analysis or regulome (ChIP, small RNA) analysis. Proprietary and third party tools are used for the analysis and handling of Next Gen sequencing data.

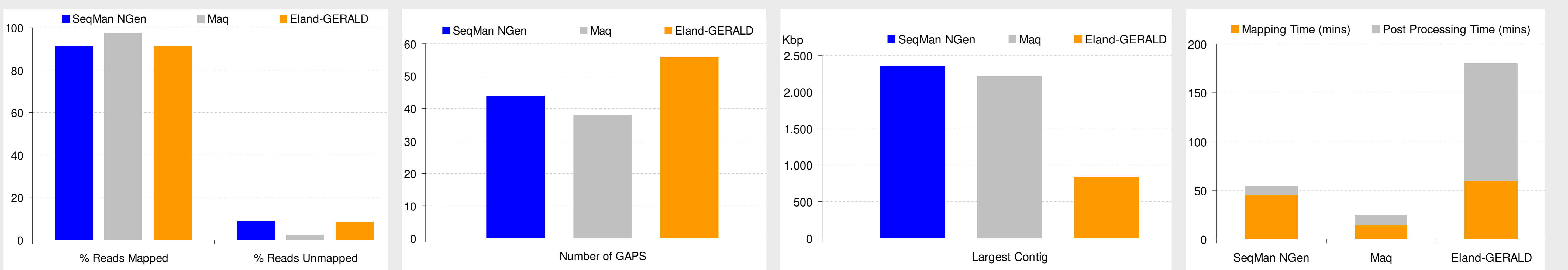
#### De Novo Assembly

Genome size: 2.5 MB  
 Sequencing: using Roche GS FLX Standard | 300,132 reads | 72 Mio bases  
 Assembly: DNASTAR SeqMan NGen™, Roche Diagnostic's GS De Novo Assembler (Newbler Assembler) & MIRA<sup>1</sup> (Mimicking Intelligent Read Assembly)



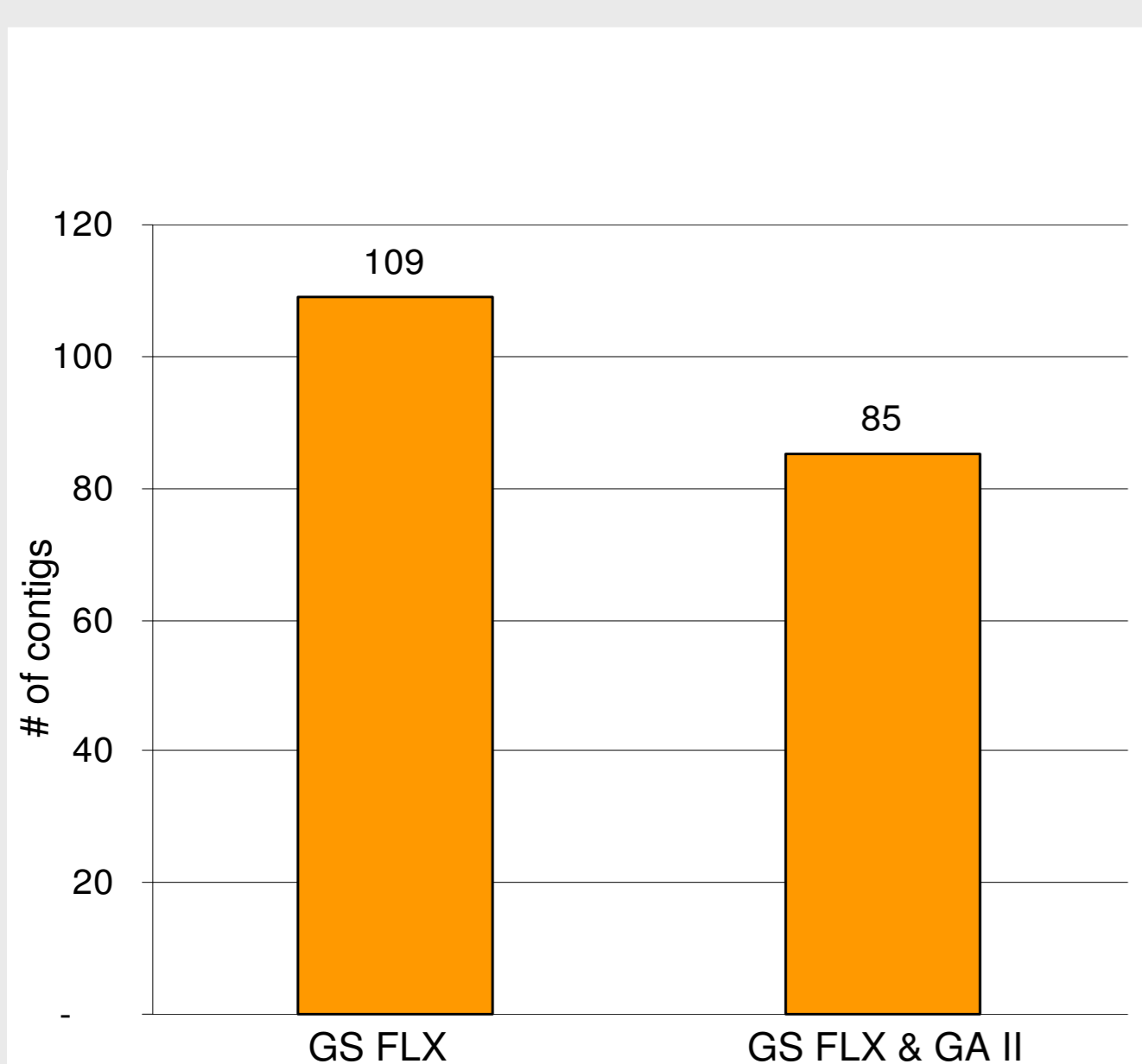
#### Resequencing / Mapping

Genome size: 4.6 MB  
 Sequencing: using Illumina Genome Analyzer II Paired-End (2\*36 bp) | 6.8 Mio reads | 247 Mio bases  
 Alignment: DNASTAR SeqMan NGen™, MAQ<sup>2</sup> (Mapping and Assembly with Quality) & Eland-GERALD

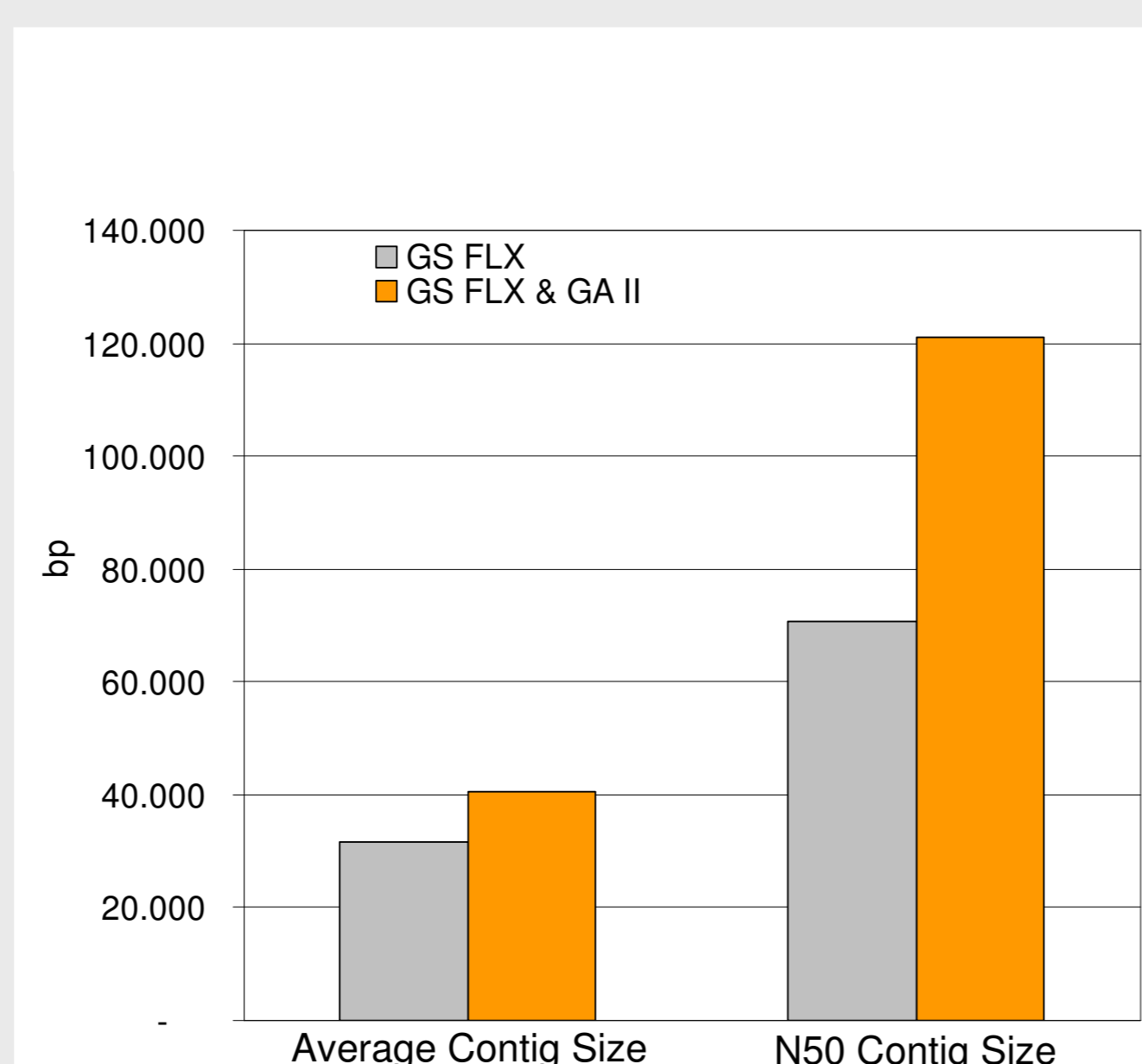


#### Hybrid Assembly

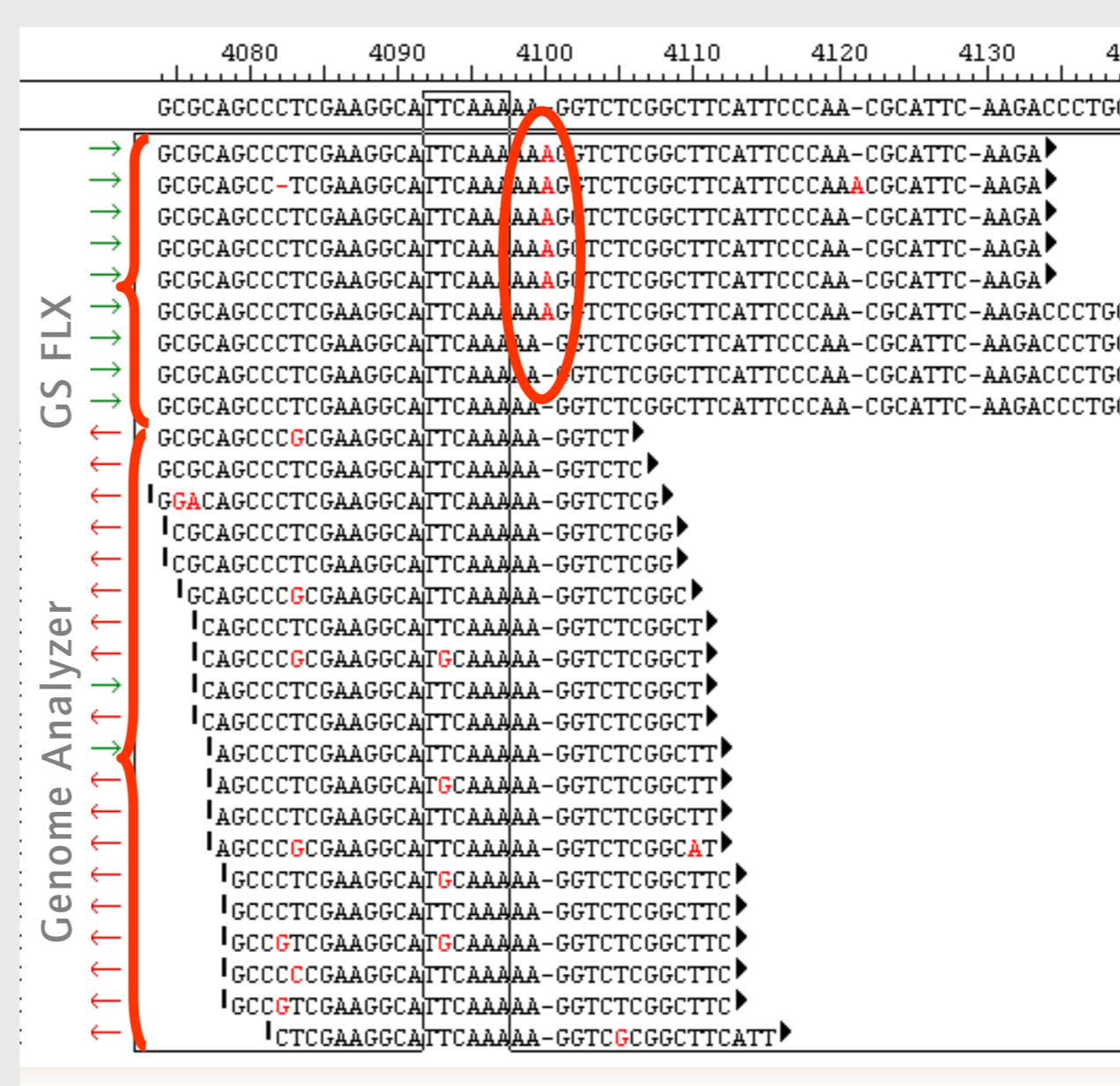
Sequencing on Roche GS FLX: 402.186 reads resp. 95.029.396 bases  
 Sequencing on Illumina GA II: 6.439.996 reads resp. 213.893.856 bases



Decrease number of contigs



Increase of average contig size and N50 contig size



Validation of homopolymers

#### Conclusion

1. Depending on the goals of the project, different sequencing technologies or combination of technologies should be applied.
2. The project goals and technologies used dictate the analysis routines involving a wide range of rapidly changing bioinformatic tools.
3. In-depth knowledge of the strength and limitations of the Next Gen technologies and of the available analysis tools is the most crucial issue in project design and analysis.