# Next Generation Sequencing

Dr. Christopher Bauser
GATC Biotech

12 May 2009

© 2007-2009

# About GATC Biotech

- ***GATC Biotech is a leading provider of DNA sequencing***
- ***services and innovative bioinformatics solutions***

- ***Founded 1990***
- ***for 10,000 customers***
  *from industry & academia*
- ***in 40 countries***
  *worldwide*
- ***100% self-financed,***
  ***independent & profitable***
- ***90 employees in 4 subsidaries***
  ***& 10 sales offices***

Company | Portfolio | References

© 2007-2009

# European wide sales structure

- *We have subsidiaries and offices in*

- ***Germany***
  ***Constance, Berlin, Munich, Stuttgart, Bielefeld***
- ***United Kingdom***
  ***London, Cambridge, Edinburgh***
- ***France***
  ***Paris, Marseille, Lyon***
- ***Sweden***
  ***Stockholm***

Company Portfolio References

© 2007-2009

# Sequencing technologies in house

Applied Biosystems
ABI 3730XL

since 1996

Roche / 454
Genome Sequencer FLX

Jan. 07 GS20
today 2 GS FLX
(Titanium chemistry)

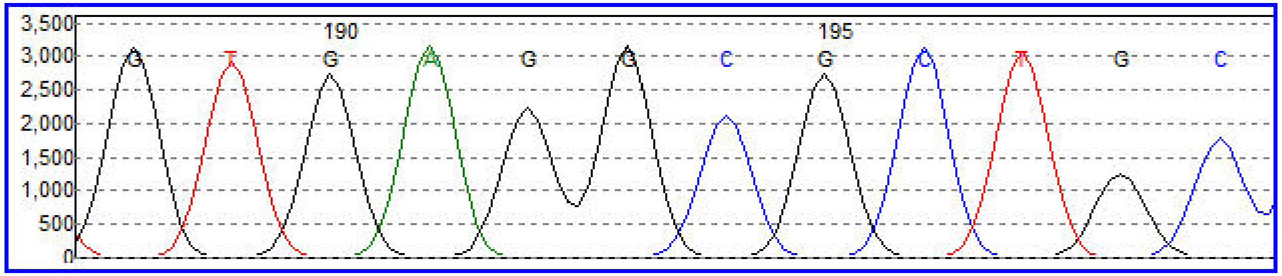
Illumina / Solexa
Genome Analyzer II

Jan. 07 GA classic
today 2 GA II
(2 x 76 bases)

© 2007-2009

# Applied Biosystems : ABI 3730

- Sanger sequencing
- The *Gold Standard* in DNA sequence analysis
- dideoxynucleotide chain termination
- read lenght up to 1,100 bases
- 1 Mb / machine / day



© 2007-2009

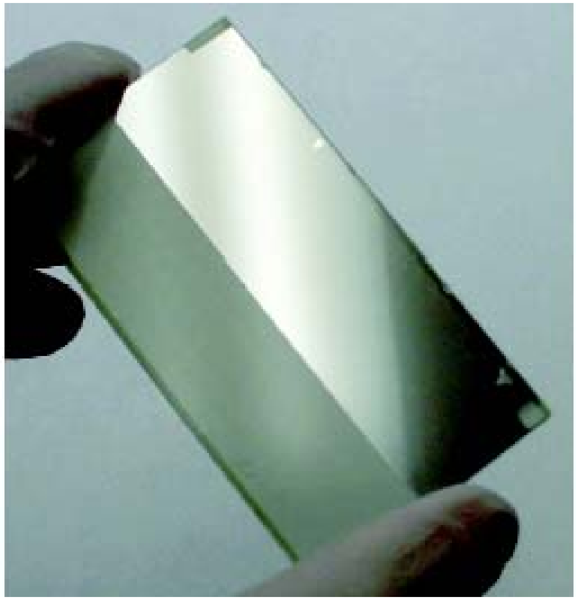
# Roche / 454: GS FLX overview

Pyrosequencing

Amplification: emulsion PCR

Standard run

- up to 400,000 reads / run
- Ø 250 bases / read
- ~ 100 Mb / run



Titanium run

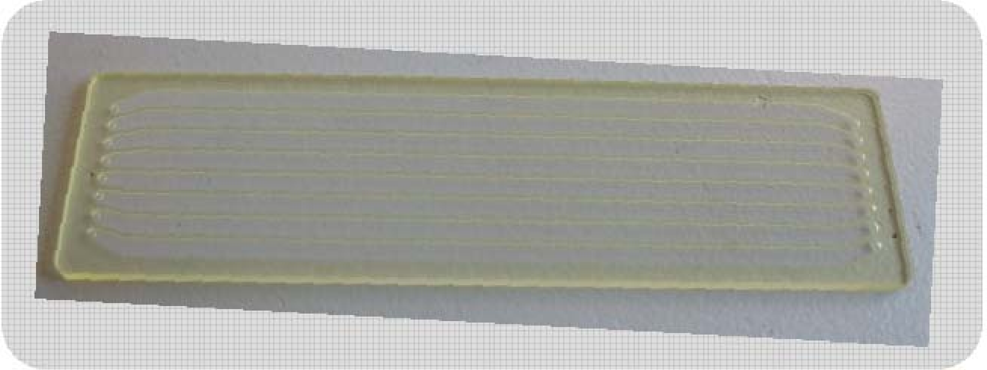
- > 1,000,000 reads / run
- Ø 350 - 450 bases / read
- ~ 400 Mb / run

© 2007-2009

# Illumina / Solexa: GAII Overview

Sequencing by synthesis

Amplification: bridging PCR

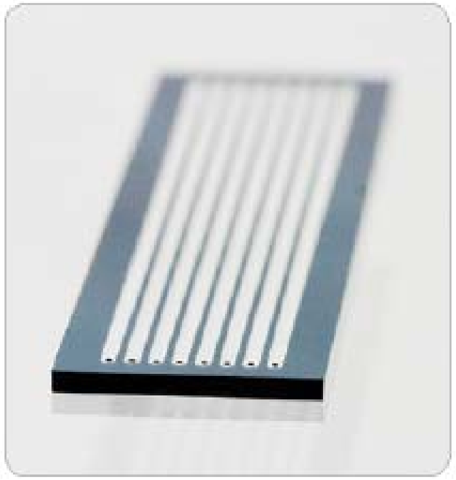


8 Lanes / run

- 40bp reads
- 76bp reads
- Up to 6Gbp / run

Paired end / mate pairs

- 200bp insert size
- 3k insert size
- 10k insert size (comming soon)



© 2007-2009

# Sequencing of...

| | |
|---|---|
| Genomes | *de novo* sequencing<br>re-sequencing |
| Transcriptomes | cDNA<br>amplicons, exons<br>SAGE / CAGE; 3’ UTR / 5’ UTR |
| Metagenomes | whole metagenome<br>prokaryotic cDNA<br>16S / 18S rRNA |
| Regulomes | smallRNA / microRNA<br>ChIP |

© 2007-2009

# Applications

| | |
|---|---|
| **Genomes** | ***de novo* sequencing**<br>re-sequencing |
| Transcriptomes | cDNA<br>amplicons, exons<br>SAGE / CAGE; 3’ UTR / 5’ UTR |
| Metagenomes | whole metagenome<br>prokaryotic cDNA<br>16S / 18S rRNA |
| Regulomes | smallRNA / microRNA<br>ChIP |

# *De novo* sequencing: Hybrid Strategy

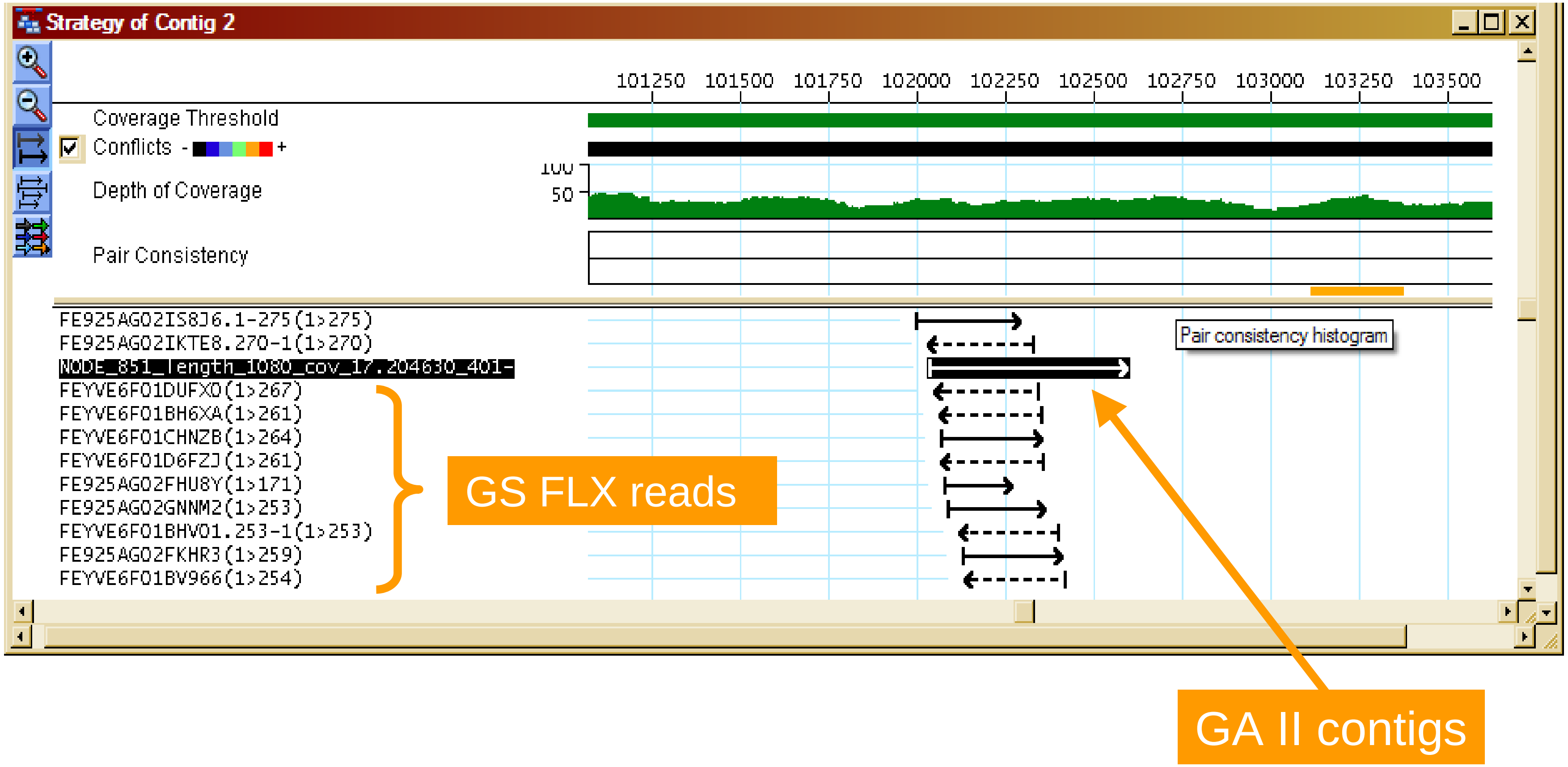
Raw data > Assembly > Alignment & homopolymers > Scaffolds & validation > Visualization

| | |
|---|---|
| 1. Raw sequence data | fasta, qual & sff |
| of GS FLX & GA II | fasta, qual & scarf |
| 2. a) *de novo* assembly GA II | fasta & qual |
| b) Hybrid assembly GA II contigs + GS FLX raw | ace |
| 3. Mapping / Alignment for Homopolymer correction | ace, fasta |
| 4. Scaffolding & assembly validation | xls, ace |
| 5. Statistics, summary | pdf, txt |

© 2007-2009

# *De novo* sequencing: Hybrid Strategy

Raw data › **Assembly** › Alignment & homopolymers › Scaffolds & validation › Visualization



```
FE925AG02IS8J6.1-275(1>275)
FE925AG02IKTE8.270-1(1>270)
NODE_851_length_1080_cov_17.204630_401-
FEYVE6F01DUFXO(1>267)
FEYVE6F01BH6XA(1>261)
FEYVE6F01CHNZB(1>264)
FEYVE6F01D6FZJ(1>261)
FE925AG02FHU8Y(1>171)
FE925AG02GNNM2(1>253)
FEYVE6F01BHVO1.253-1(1>253)
FE925AG02FKHR3(1>259)
FEYVE6F01BV966(1>254)
```

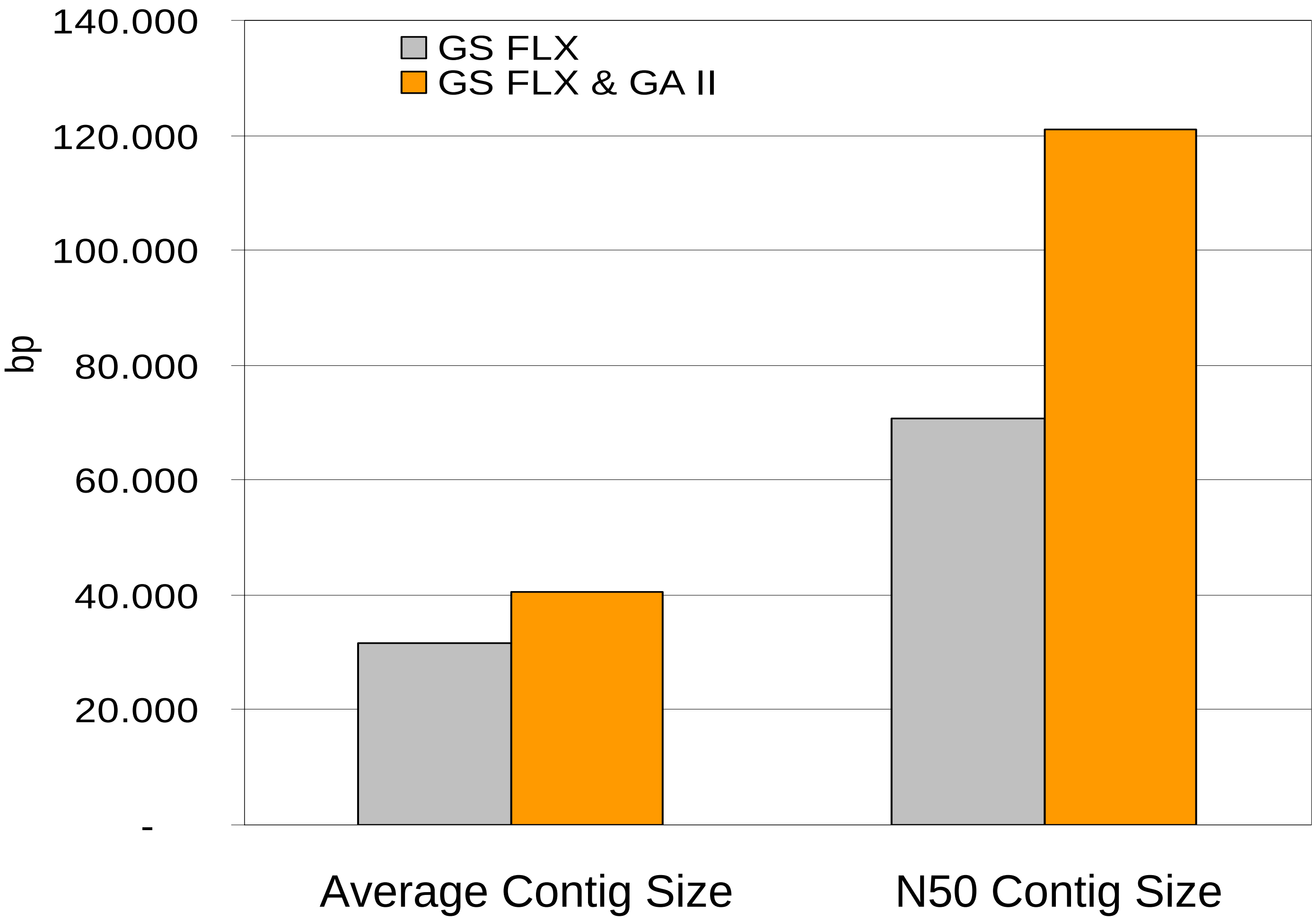
GS FLX reads

GA II contigs

© 2007-2009

# Hybrid Strategy: contig size

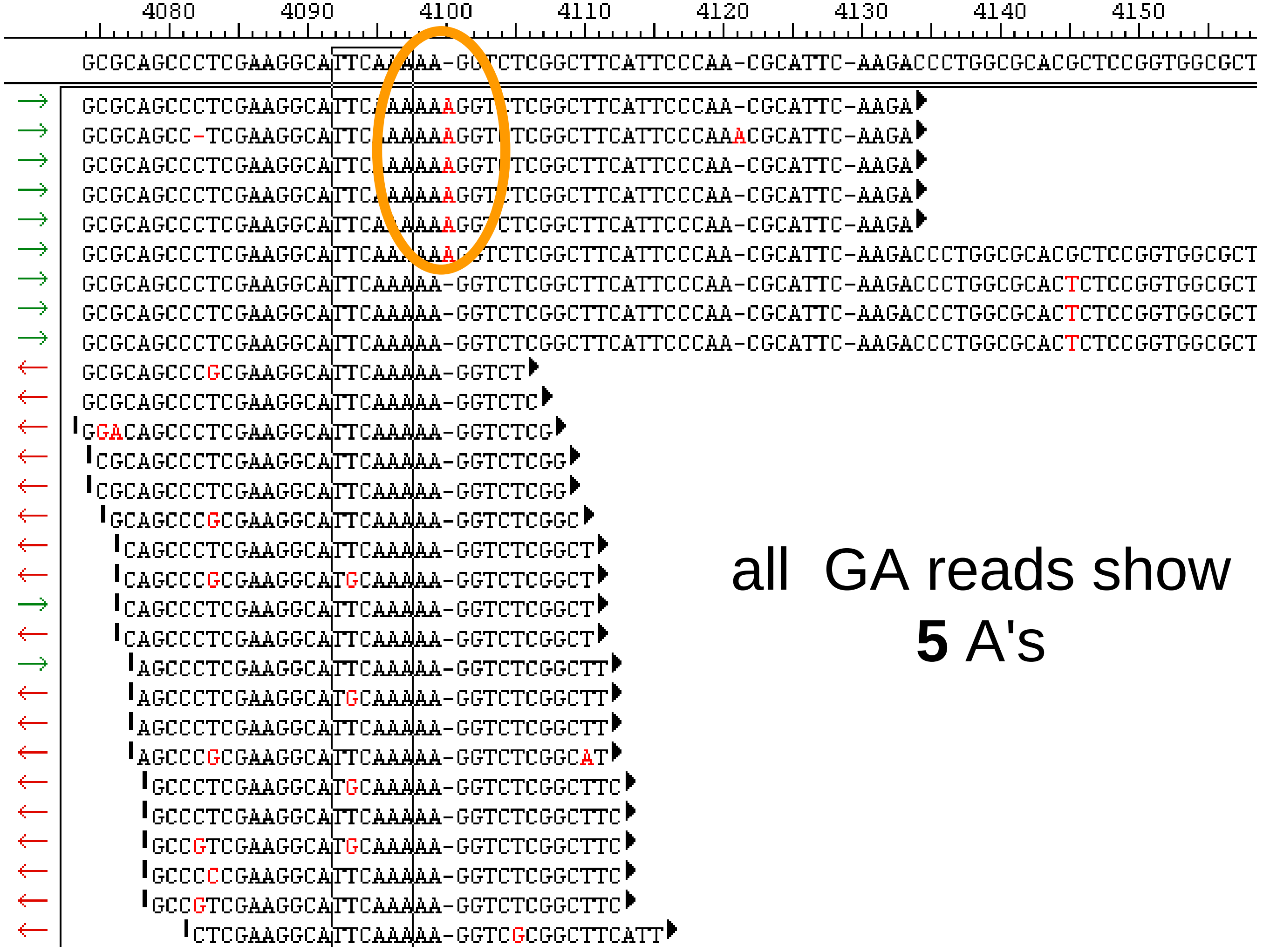
Raw data > Assembly > Alignment & homopolymers > Scaffolds & validation > Visualization



© 2007-2009

# Hybrid Strategy: homopolymers

Raw data > Assembly > **Alignment & homopolymers** > Scaffolds & validation > Visualization



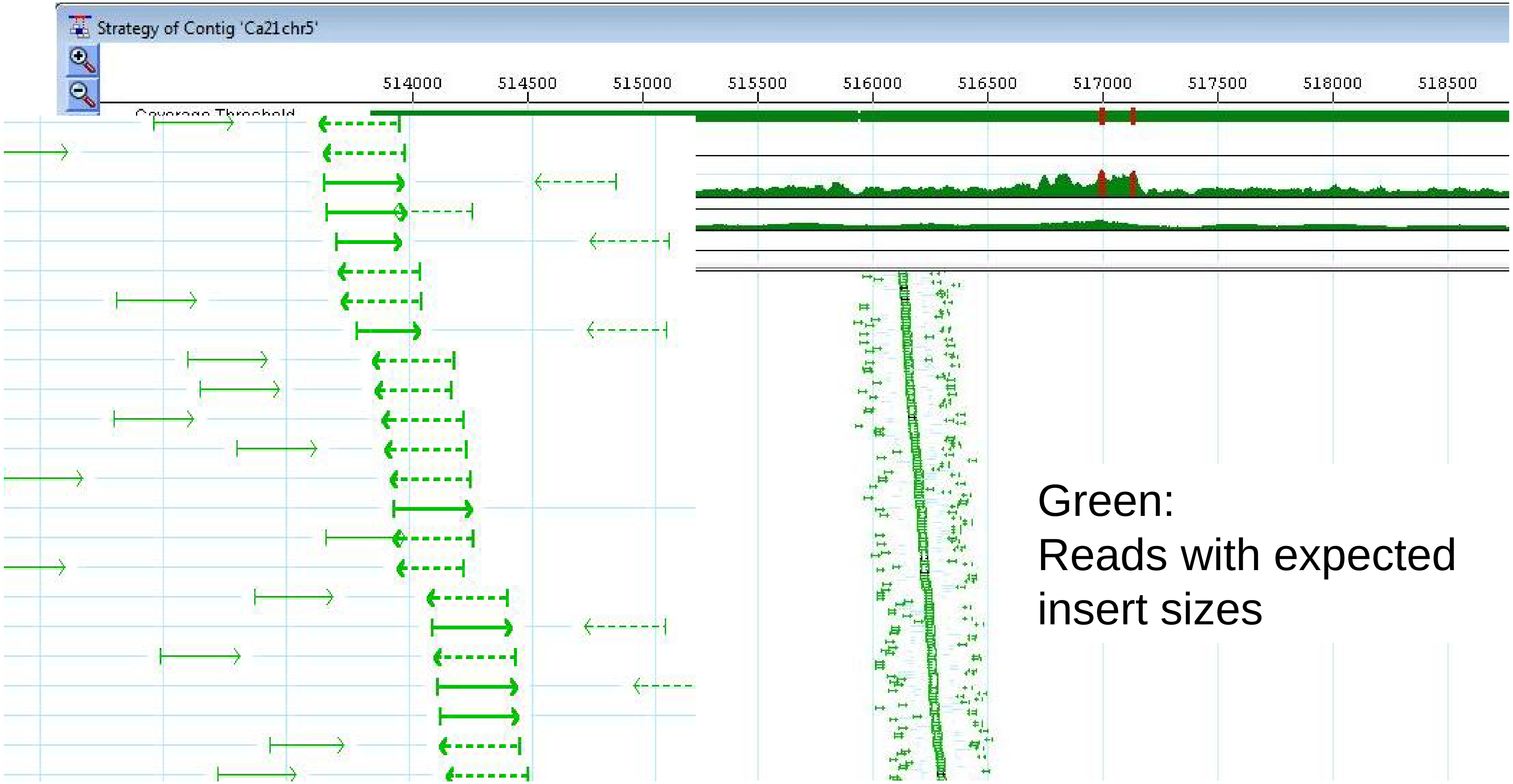
```
       4080      4090      4100      4110      4120      4130      4140      4150
  GCGCAGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCTTCATTCCCAA-CGCATTC-AAGACCCTGGCGCACGCTCCGGTGGCGCT
→ GCGCAGCCCTCGAAGGCATTCAAAAAAGGTCTCGGCTTCATTCCCAA-CGCATTC-AAGA
→ GCGCAGCC-TCGAAGGCATTCAAAAAAGGTCTCGGCTTCATTCCCAAACGCATTC-AAGA
→ GCGCAGCCCTCGAAGGCATTCAAAAAAGGTCTCGGCTTCATTCCCAA-CGCATTC-AAGA
→ GCGCAGCCCTCGAAGGCATTCAAAAAAGGTCTCGGCTTCATTCCCAA-CGCATTC-AAGA
→ GCGCAGCCCTCGAAGGCATTCAAAAAAGGTCTCGGCTTCATTCCCAA-CGCATTC-AAGA
→ GCGCAGCCCTCGAAGGCATTCAAAAAAGGTCTCGGCTTCATTCCCAA-CGCATTC-AAGACCCTGGCGCACGCTCCGGTGGCGCT
→ GCGCAGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCTTCATTCCCAA-CGCATTC-AAGACCCTGGCGCACTCTCCGGTGGCGCT
→ GCGCAGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCTTCATTCCCAA-CGCATTC-AAGACCCTGGCGCACTCTCCGGTGGCGCT
→ GCGCAGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCTTCATTCCCAA-CGCATTC-AAGACCCTGGCGCACTCTCCGGTGGCGCT
← GCGCAGCCCGCGAAGGCATTCAAAAA-GGTCT
← GCGCAGCCCTCGAAGGCATTCAAAAA-GGTCTC
← GGACAGCCCTCGAAGGCATTCAAAAA-GGTCTCG
←  CGCAGCCCTCGAAGGCATTCAAAAA-GGTCTCGG
←  CGCAGCCCTCGAAGGCATTCAAAAA-GGTCTCGG
←   GCAGCCCGCGAAGGCATTCAAAAA-GGTCTCGGC
←    CAGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCT
←    CAGCCCGCGAAGGCATGCAAAAA-GGTCTCGGCT
→    CAGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCT
←    CAGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCT
→     AGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCTT
←     AGCCCTCGAAGGCATGCAAAAA-GGTCTCGGCTT
←     AGCCCTCGAAGGCATTCAAAAA-GGTCTCGGCTT
←     AGCCCGCGAAGGCATTCAAAAA-GGTCTCGGCAT
←      GCCCTCGAAGGCATGCAAAAA-GGTCTCGGCTTC
←      GCCCTCGAAGGCATTCAAAAA-GGTCTCGGCTTC
←      GCCGTCGAAGGCATGCAAAAA-GGTCTCGGCTTC
←      GCCCCCGAAGGCATTCAAAAA-GGTCTCGGCTTC
←      GCCGTCGAAGGCATTCAAAAA-GGTCTCGGCTTC
←         CTCGAAGGCATTCAAAAA-GGTCGCGGCTTCATT
```

all GA reads show **5** A's

© 2007-2009

# Hybrid Strategy: scaffolding

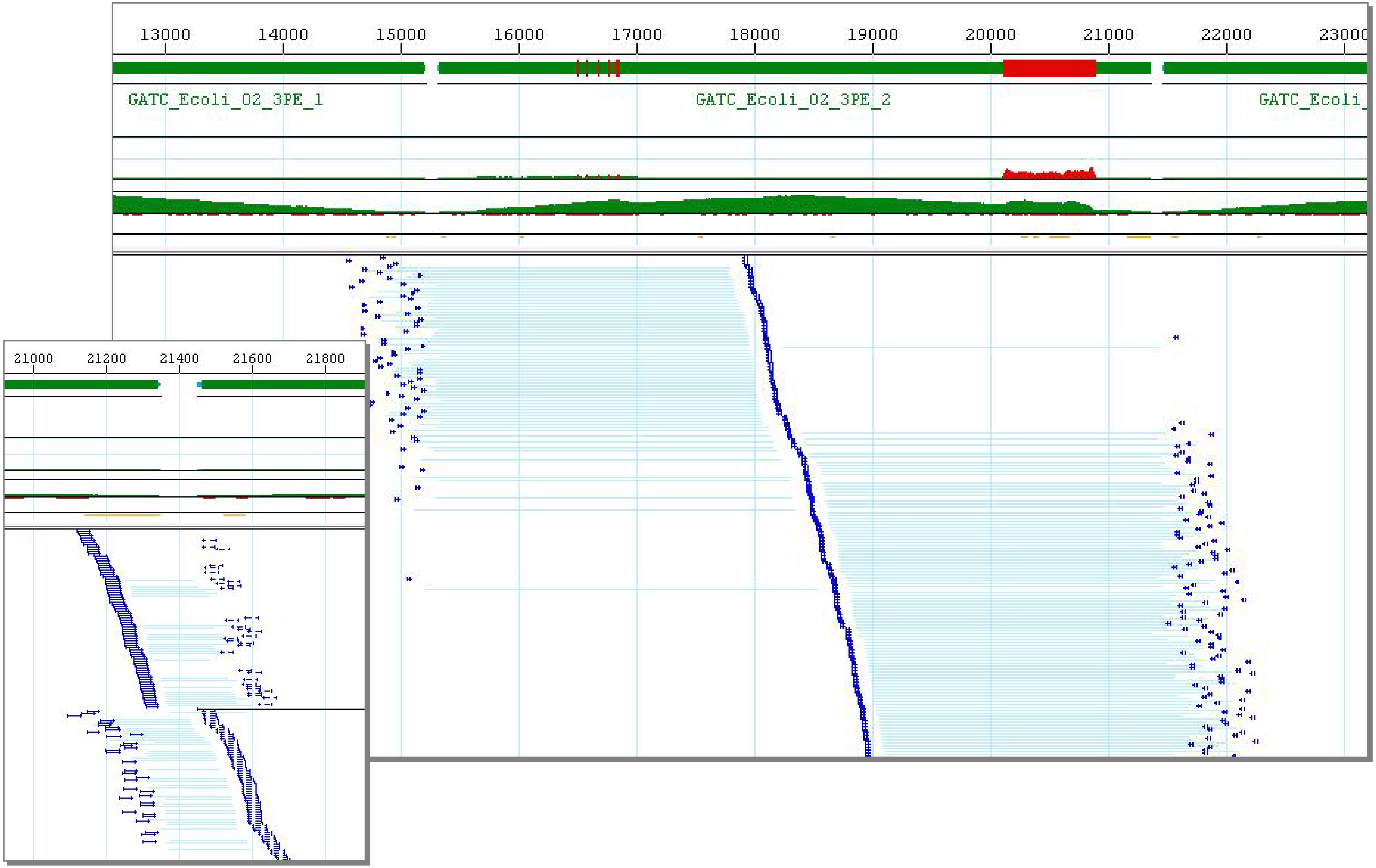
Raw data > Assembly > Alignment & homopolymers > **Scaffolds & validation** > Visualization



Green:
Reads with expected insert sizes

© 2007-2009

# Hybrid Strategy: scaffolding



© 2007-2009

# Applications

**Genomes**

*de novo* sequencing
**re-sequencing**

Transcriptomes

cDNA
amplicons, exons
SAGE / CAGE; 3’ UTR / 5’ UTR

Metagenomes

whole metagenome
prokaryotic cDNA
16S / 18S rRNA

Regulomes

smallRNA / microRNA
ChIP

© 2007-2009

# Re-Sequencing: GA II

Raw data > Alignment / Mapping & SNP > InDel > Visualization

| | |
|---|---|
| 1. Raw sequence data | fasta, qual & scarf |
| 2. Alignment / Mapping & SNP | txt, ace, gff & xls |
| 3. InDel table | txt / xls |
| 4. Alignment report | txt / xls & gff (step 2) |
| 5. Statistics, summary | pdf, txt |

© 2007-2009

# Re-Sequencing: GA II

Raw data > Alignment / Mapping & SNP > InDel > Visualization

## 1. Raw sequence data: fasta, qual & scarf

**Reads in fasta format** (GATC_sequence.fas)
>**GATC_8_1_879_612**
TAAAAGGATTTTATTTTCGTTGTTGAAAGTGCATGT
>**GATC_8_1_880_245**
TATGTTGGTTGGCTTCGGTAAGGCATCCTACCTTGT

**Read Qualities in fasta Format** (GATC_sequence.qual)
>**GATC_8_1_879_612**
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 36
40 40 40 40 40
>**GATC_8_1_880_245**
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 33 35 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40

**Solexa Scarf Format** (GATC_sequence.txt)
**GATC:8:1:879:612**:TAAAAGGATTTTATTTTCGTTGTTGAAAGTGCATGT:40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 36 40 40 40 40 40
**GATC:8:1:880:245**:TATGTTGGTTGGCTTCGGTAAGGCATCCTACCTTGT:40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 33 35 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40

© 2007-2009

# Re-Sequencing: GA II

Raw data | Alignment / Mapping & SNP | InDel | Visualization

## SNP

```
>NC_000913_inRef_1    TAGGTGATGG TATGCGCACC A TGCG
<NC_000913_inRef_2    TAGGTGATGG TATGCGCACC A TGCGTG
<NC_000913_inRef_3    TAGGTGATGG TATGCGCACC A TGCGTGGG
>NC_000913_inRef_4    TAGGTGATGG TATGCGCACC A TGCGTGGGA
 NC_000913_ref        TAGGTGATGG TATGCGCACC T TGCGTGGGA ATTCTTTGC
<NC_000913_inRef_5       GTGATGG TATGCGCACC A TGCGTGGGA ATT
>NC_000913_inRef_6          ATGG TATGCGCACC A TGCGTGGGA ATTCTT
<NC_000913_inRef_7             G TATGCGCACC A TGCGTGGGA ATTCTTTGC
<NC_000913_inRef_8                 TGCGCACC A TGCGTGGGA ATTCTTTGC
```

© 2007-2009

# Re-Sequencing: GA II

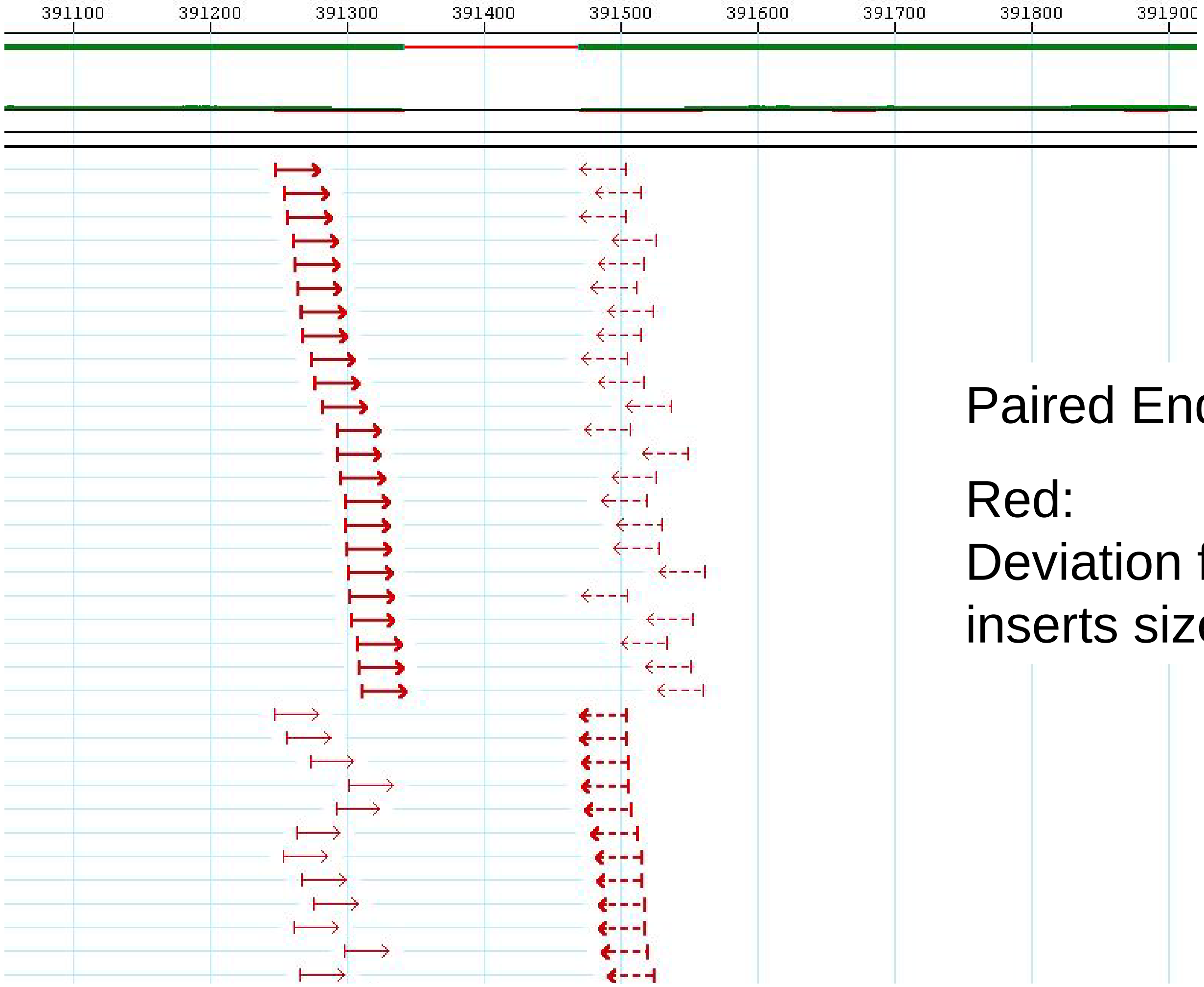
Raw data > **Alignment / Mapping & SNP** > InDel > Visualization

| Position | Coverage | Reference | Consensus | Qual(G) | Qual(A) | Qual(T) | Qual(C) | Qual(N) | Base list | Qual% |
|---|---|---|---|---|---|---|---|---|---|---|
| 41649 | 19 | C | T | 0 | 0 | 313 | 257 | 9 | CCCCCCCTT | 0,54 |
| 41726 | 11 | A | G | 224 | 102 | 0 | 0 | 12 | AAAGGGGG | 0,68 |
| 42339 | 21 | A | C | 0 | 189 | 0 | 352 | 35 | AAAAACCCC | 0,63 |
| 43461 | 24 | C | T | 0 | 0 | 253 | 180 | 29 | CCCCCCTTT | 0,57 |
| 43896 | 22 | C | T | 0 | 0 | 345 | 273 | 24 | CCCCCCCTT | 0,56 |
| 45140 | 21 | A | G | 324 | 206 | 0 | 0 | 22 | AAAAAAAAC | 0,6 |
| 45369 | 11 | T | G | 156 | 0 | 80 | 0 | 11 | GGGGGTTgt | 0,66 |
| 45559 | 18 | C | A | 0 | 235 | 0 | 212 | 32 | AAAAAAACC | 0,49 |
| 45636 | 24 | C | T | 0 | 0 | 266 | 260 | 18 | CCCCCCCCC | 0,5 |
| 45748 | 24 | G | A | 217 | 378 | 0 | 0 | 25 | AAAAAAAAA | 0,62 |
| 51426 | 15 | A | G | 201 | 181 | 0 | 0 | 9 | AAAAAAAGC | 0,51 |
| 72440 | 17 | T | C | 0 | 0 | 0 | 547 | 16 | CCCCCCCCC | 1 |
| 86634 | 4 | G | C | 21 | 0 | 0 | 26 | 8 | CGgg | 0,47 |
| 96307 | 26 | A | C | 0 | 0 | 0 | 715 | 17 | CCCCCCCCC | 1 |
| 96350 | 6 | T | C | 0 | 0 | 0 | 153 | 0 | CCCCCC | 1 |
| 96353 | 4 | T | G | 80 | 0 | 0 | 0 | 12 | GGgg | 1 |
| 96393 | 6 | A | C | 0 | 0 | 0 | 169 | 7 | CCCCCc | 1 |
| 96408 | 5 | G | C | 0 | 0 | 0 | 17 | 2 | Ccccc | 1 |
| 102433 | 11 | T | G | 98 | 0 | 80 | 0 | 21 | GGGGTTttttt | 0,49 |
| 113988 | 14 | T | C | 0 | 0 | 117 | 223 | 21 | CCCCCCCTT | 0,62 |
| 114681 | 25 | C | T | 0 | 0 | 484 | 322 | 10 | CCCCCCCCC | 0,6 |
| 151966 | 26 | A | G | 999 | 0 | 0 | 0 | 0 | GGGGGGGG | 1 |

© 2007-2009

# Re-Sequencing GA II: mate pairs / paired end

Raw data > Alignment / Mapping & SNP > InDel > Visualization



Paired End Sequence Reads

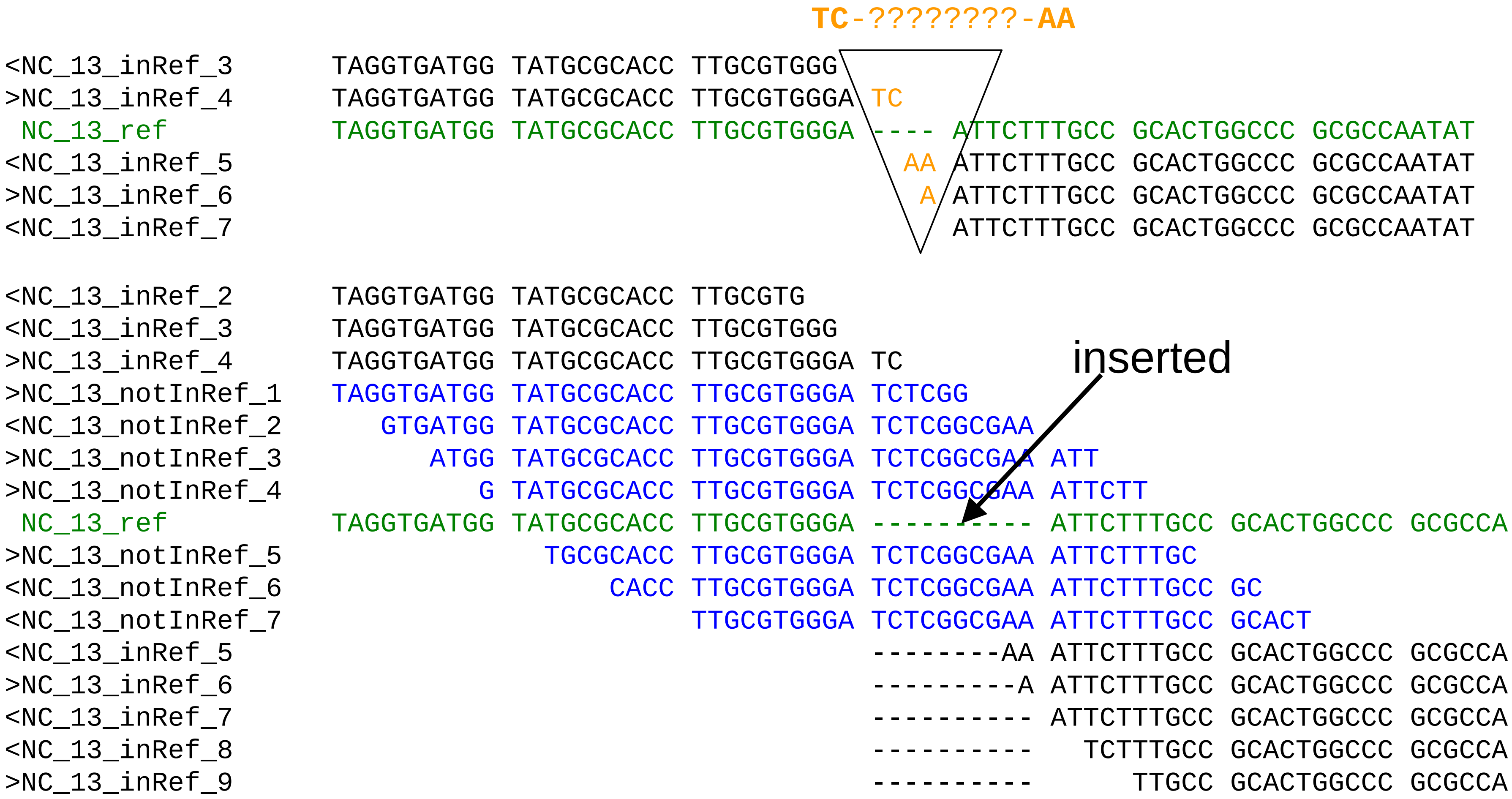
Red:
Deviation from expected inserts sizes

© 2007-2009

# Re-Sequencing: insertion

Raw data > Alignment / Mapping & SNP > **InDel** > Visualization

## Subsequent analysis with proprietary GATC tools



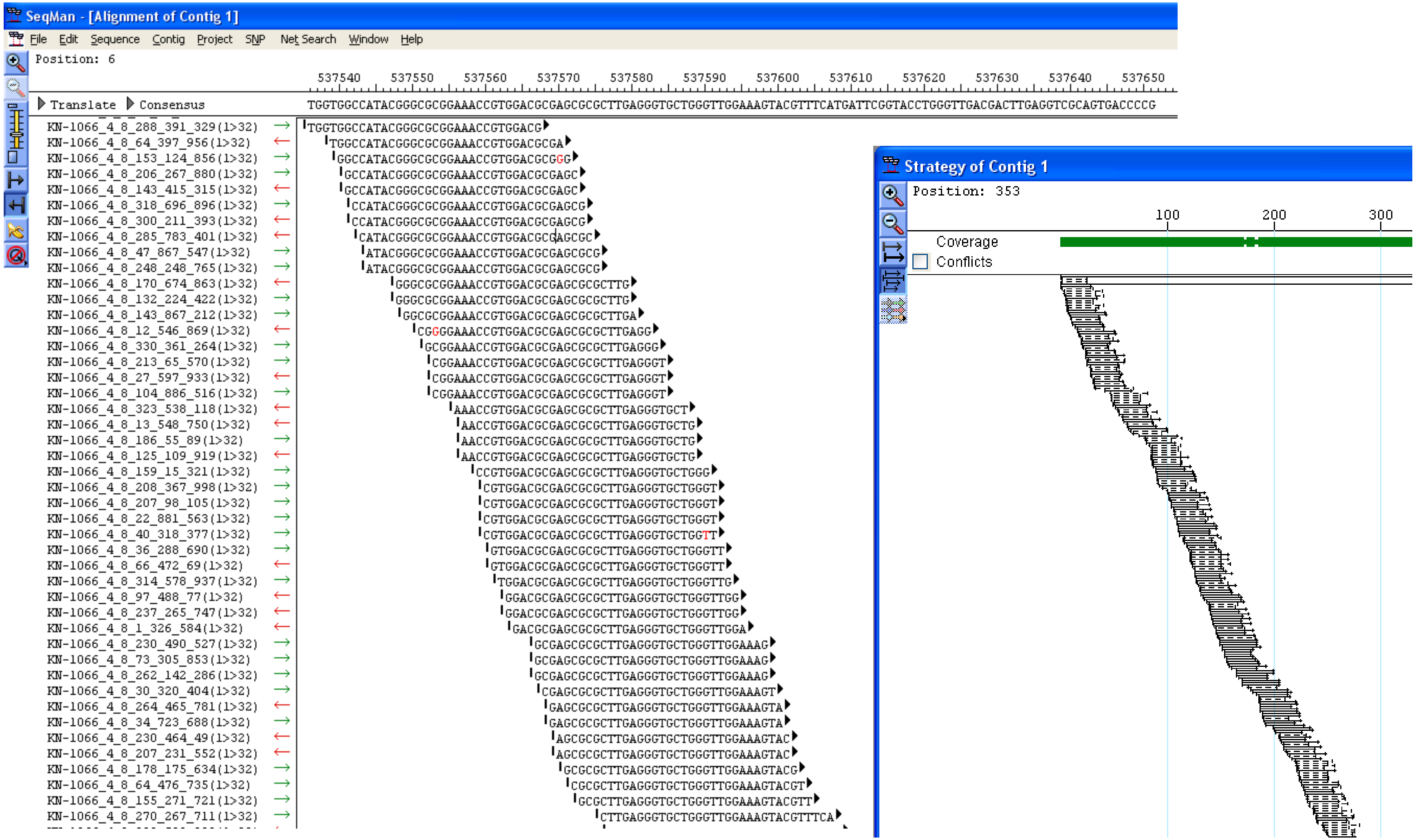
```
                                                  TC-????????-AA

<NC_13_inRef_3      TAGGTGATGG TATGCGCACC TTGCGTGGG
>NC_13_inRef_4      TAGGTGATGG TATGCGCACC TTGCGTGGGA TC
 NC_13_ref          TAGGTGATGG TATGCGCACC TTGCGTGGGA ---- ATTCTTTGCC GCACTGGCCC GCGCCAATAT
<NC_13_inRef_5                                         AA ATTCTTTGCC GCACTGGCCC GCGCCAATAT
>NC_13_inRef_6                                          A ATTCTTTGCC GCACTGGCCC GCGCCAATAT
<NC_13_inRef_7                                            ATTCTTTGCC GCACTGGCCC GCGCCAATAT

<NC_13_inRef_2      TAGGTGATGG TATGCGCACC TTGCGTG
<NC_13_inRef_3      TAGGTGATGG TATGCGCACC TTGCGTGGG
>NC_13_inRef_4      TAGGTGATGG TATGCGCACC TTGCGTGGGA TC
>NC_13_notInRef_1   TAGGTGATGG TATGCGCACC TTGCGTGGGA TCTCGG
<NC_13_notInRef_2      GTGATGG TATGCGCACC TTGCGTGGGA TCTCGGCGAA
>NC_13_notInRef_3         ATGG TATGCGCACC TTGCGTGGGA TCTCGGCGAA ATT
>NC_13_notInRef_4            G TATGCGCACC TTGCGTGGGA TCTCGGCGAA ATTCTT
 NC_13_ref          TAGGTGATGG TATGCGCACC TTGCGTGGGA ---------- ATTCTTTGCC GCACTGGCCC GCGCCA
>NC_13_notInRef_5                TGCGCACC TTGCGTGGGA TCTCGGCGAA ATTCTTTGC
<NC_13_notInRef_6                    CACC TTGCGTGGGA TCTCGGCGAA ATTCTTTGCC GC
<NC_13_notInRef_7                         TTGCGTGGGA TCTCGGCGAA ATTCTTTGCC GCACT
<NC_13_inRef_5                                       --------AA ATTCTTTGCC GCACTGGCCC GCGCCA
>NC_13_inRef_6                                       ---------A ATTCTTTGCC GCACTGGCCC GCGCCA
<NC_13_inRef_7                                       ---------- ATTCTTTGCC GCACTGGCCC GCGCCA
<NC_13_inRef_8                                       ----------   TCTTTGCC GCACTGGCCC GCGCCA
>NC_13_inRef_9                                       ----------      TTGCC GCACTGGCCC GCGCCA
```

inserted

© 2007-2009

# Re-Sequencing: GA II

Raw data > Alignment / Mapping & SNP > InDel > **Visualization**



```
Position: 6
                                   537540    537550    537560    537570    537580    537590    537600    537610    537620    537630    537640    537650
Translate   Consensus              TGGTGGCCATACGGGCGCGGAAACCGTGGACGCGAGCGCGCTTGAGGGTGCTGGGTTGGAAAGTACGTTTCATGATTCGGTACCTGGGTTGACGACTTGAGGTCGCAGTGACCCCG
KN-1066_4_8_288_391_329(1>32) →  TGGTGGCCATACGGGCGCGGAAACCGTGGACG
KN-1066_4_8_64_397_956(1>32)  ←     TGGCCATACGGGCGCGGAAACCGTGGACGCGA
KN-1066_4_8_153_124_856(1>32) →      GGCCATACGGGCGCGGAAACCGTGGACGCGGG
KN-1066_4_8_206_267_880(1>32) →       GCCATACGGGCGCGGAAACCGTGGACGCGAGC
KN-1066_4_8_143_415_315(1>32) ←       GCCATACGGGCGCGGAAACCGTGGACGCGAGC
KN-1066_4_8_318_696_896(1>32) →        CCATACGGGCGCGGAAACCGTGGACGCGAGCG
KN-1066_4_8_300_211_393(1>32) ←        CCATACGGGCGCGGAAACCGTGGACGCGAGCG
KN-1066_4_8_285_783_401(1>32) ←         CATACGGGCGCGGAAACCGTGGACGCGAGCGC
KN-1066_4_8_47_867_547(1>32)  →          ATACGGGCGCGGAAACCGTGGACGCGAGCGCG
KN-1066_4_8_248_248_765(1>32) →          ATACGGGCGCGGAAACCGTGGACGCGAGCGCG
KN-1066_4_8_170_674_863(1>32) ←              GGGCGCGGAAACCGTGGACGCGAGCGCGCTTG
KN-1066_4_8_132_224_422(1>32) →              GGGCGCGGAAACCGTGGACGCGAGCGCGCTTG
KN-1066_4_8_143_867_212(1>32) →               GGCGCGGAAACCGTGGACGCGAGCGCGCTTGA
KN-1066_4_8_12_546_869(1>32)  ←                 CGGGGAAACCGTGGACGCGAGCGCGCTTGAGG
KN-1066_4_8_330_361_264(1>32) →                  GCGGAAACCGTGGACGCGAGCGCGCTTGAGGG
KN-1066_4_8_213_65_570(1>32)  →                   CGGAAACCGTGGACGCGAGCGCGCTTGAGGGT
KN-1066_4_8_27_597_933(1>32)  ←                   CGGAAACCGTGGACGCGAGCGCGCTTGAGGGT
KN-1066_4_8_104_886_516(1>32) →                   CGGAAACCGTGGACGCGAGCGCGCTTGAGGGT
KN-1066_4_8_323_538_118(1>32) ←                      AAACCGTGGACGCGAGCGCGCTTGAGGGTGCT
KN-1066_4_8_13_548_750(1>32)  ←                       AACCGTGGACGCGAGCGCGCTTGAGGGTGCTG
KN-1066_4_8_186_55_89(1>32)   →                       AACCGTGGACGCGAGCGCGCTTGAGGGTGCTG
KN-1066_4_8_125_109_919(1>32) ←                       AACCGTGGACGCGAGCGCGCTTGAGGGTGCTG
KN-1066_4_8_159_15_321(1>32)  →                         CCGTGGACGCGAGCGCGCTTGAGGGTGCTGGG
KN-1066_4_8_208_367_998(1>32) →                          CGTGGACGCGAGCGCGCTTGAGGGTGCTGGGT
KN-1066_4_8_207_98_105(1>32)  →                          CGTGGACGCGAGCGCGCTTGAGGGTGCTGGGT
KN-1066_4_8_22_881_563(1>32)  →                          CGTGGACGCGAGCGCGCTTGAGGGTGCTGGGT
KN-1066_4_8_40_318_377(1>32)  →                          CGTGGACGCGAGCGCGCTTGAGGGTGCTGGTT
KN-1066_4_8_36_288_690(1>32)  →                           GTGGACGCGAGCGCGCTTGAGGGTGCTGGGTT
KN-1066_4_8_66_472_69(1>32)   ←                           GTGGACGCGAGCGCGCTTGAGGGTGCTGGGTT
KN-1066_4_8_314_578_937(1>32) →                            TGGACGCGAGCGCGCTTGAGGGTGCTGGGTTG
KN-1066_4_8_97_488_77(1>32)   ←                             GGACGCGAGCGCGCTTGAGGGTGCTGGGTTGG
KN-1066_4_8_237_265_747(1>32) ←                             GGACGCGAGCGCGCTTGAGGGTGCTGGGTTGG
KN-1066_4_8_1_326_584(1>32)   ←                              GACGCGAGCGCGCTTGAGGGTGCTGGGTTGGA
KN-1066_4_8_230_490_527(1>32) →                                 GCGAGCGCGCTTGAGGGTGCTGGGTTGGAAAG
KN-1066_4_8_73_305_853(1>32)  →                                 GCGAGCGCGCTTGAGGGTGCTGGGTTGGAAAG
KN-1066_4_8_262_142_286(1>32) →                                 GCGAGCGCGCTTGAGGGTGCTGGGTTGGAAAG
KN-1066_4_8_30_320_404(1>32)  →                                  CGAGCGCGCTTGAGGGTGCTGGGTTGGAAAGT
KN-1066_4_8_264_465_781(1>32) ←                                   GAGCGCGCTTGAGGGTGCTGGGTTGGAAAGTA
KN-1066_4_8_34_723_688(1>32)  →                                   GAGCGCGCTTGAGGGTGCTGGGTTGGAAAGTA
KN-1066_4_8_230_464_49(1>32)  ←                                    AGCGCGCTTGAGGGTGCTGGGTTGGAAAGTAC
KN-1066_4_8_207_231_552(1>32) ←                                    AGCGCGCTTGAGGGTGCTGGGTTGGAAAGTAC
KN-1066_4_8_178_175_634(1>32) →                                     GCGCGCTTGAGGGTGCTGGGTTGGAAAGTACG
KN-1066_4_8_64_476_735(1>32)  →                                      CGCGCTTGAGGGTGCTGGGTTGGAAAGTACGT
KN-1066_4_8_155_271_721(1>32) →                                       GCGCTTGAGGGTGCTGGGTTGGAAAGTACGTT
KN-1066_4_8_270_267_711(1>32) →                                          CTTGAGGGTGCTGGGTTGGAAAGTACGTTTCA
```

```
Position: 353
Coverage
☐ Conflicts
```

© 2007-2009

# Applications

Genomes

*de novo* sequencing
re-sequencing

**Transcriptomes**

cDNA
amplicons, exons
SAGE / CAGE; 3’ UTR / 5’ UTR

Metagenomes

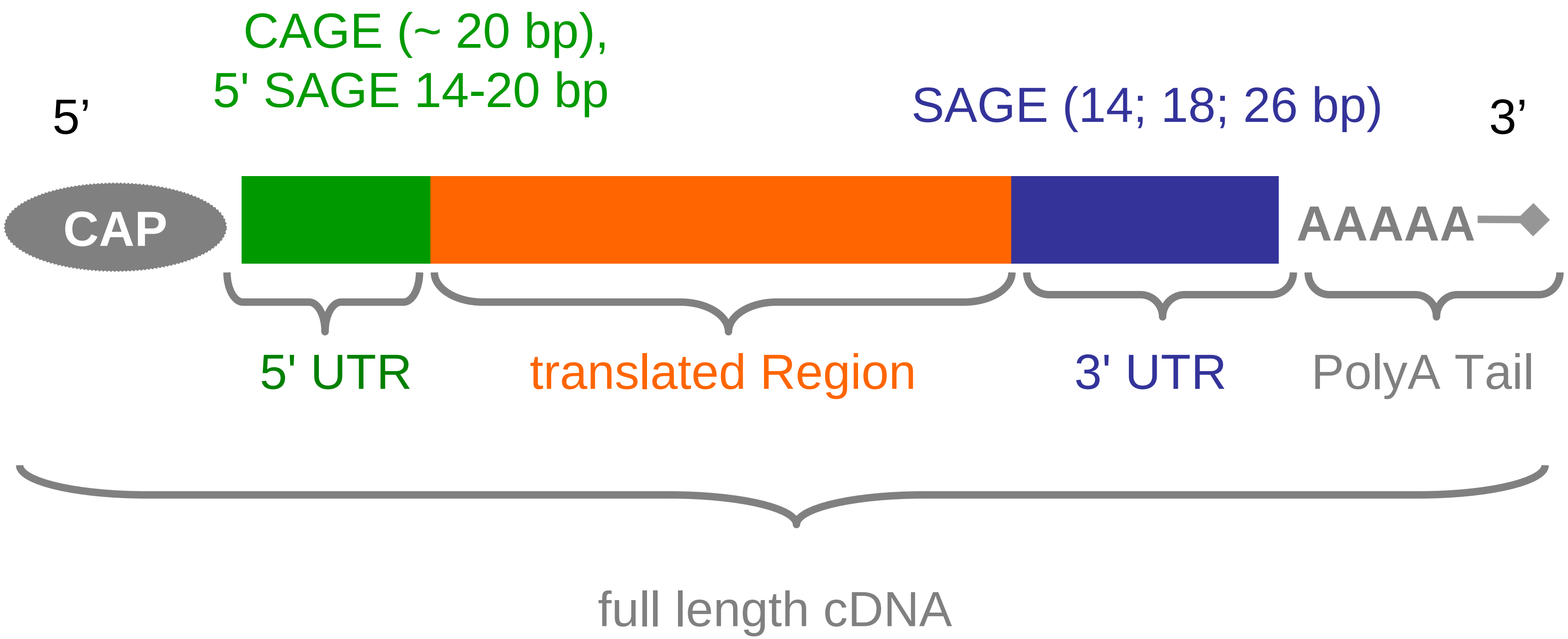
whole metagenome
prokaryotic cDNA
16S / 18S rRNA

Regulomes

smallRNA / microRNA
ChIP

© 2007-2009

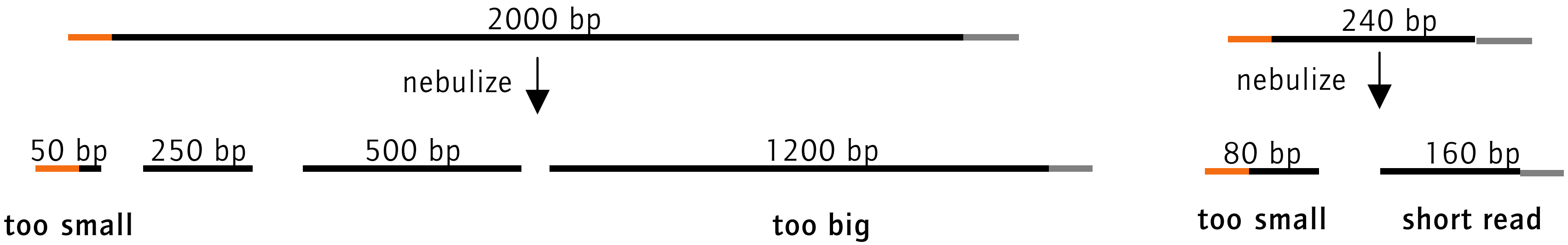
# Transcriptome



© 2007-2009

# Transcriptome: cDNA

without coligation:
smaller fragments will be lost; lots of short reads



2000 bp

nebulize

50 bp 250 bp 500 bp 1200 bp

too small

too big

240 bp

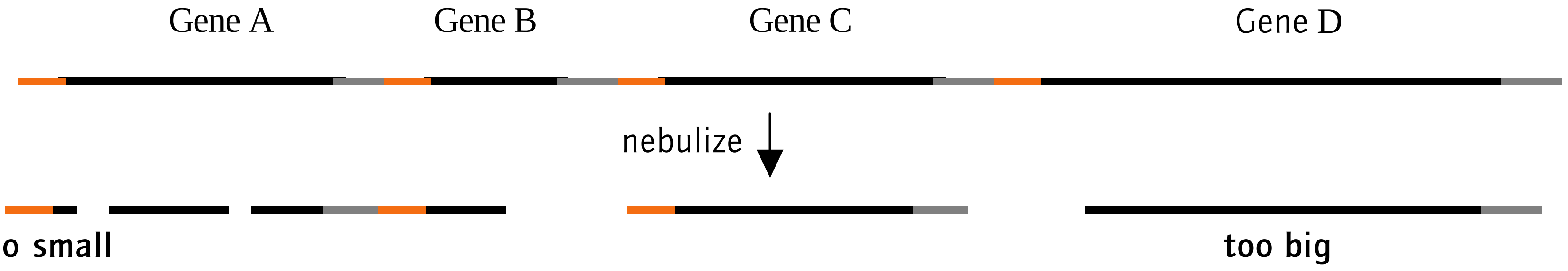
nebulize

80 bp 160 bp

too small

short read

with coligation: all fragments will be sequenced



Gene A Gene B Gene C Gene D

nebulize

too small

too big

© 2007-2009

# Transcriptome: cDNA

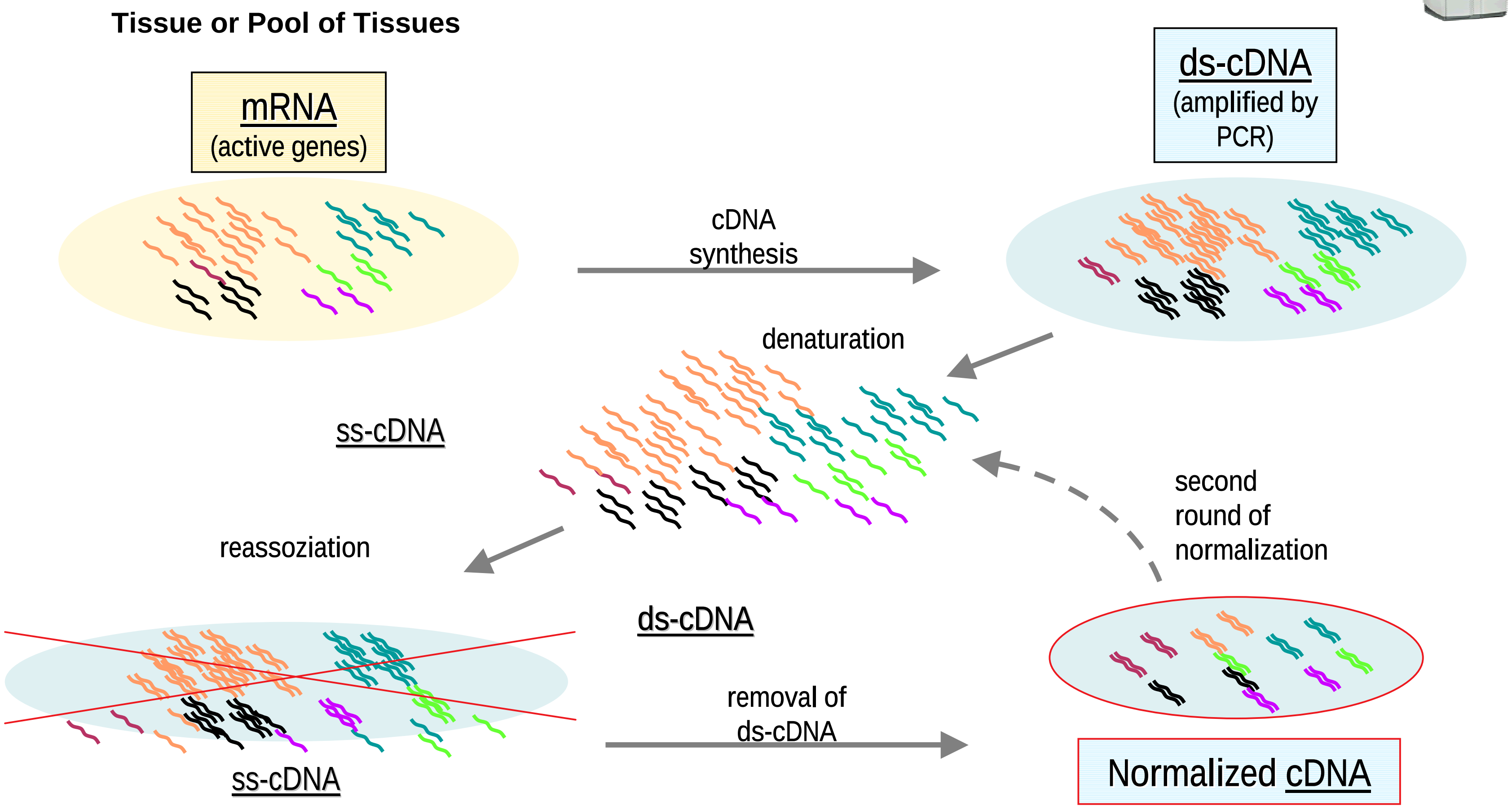
cDNA Lib > Raw Data > Clustering > GA Mapping > Visualization

| | | |
|---|---|---|
| 1. | cDNA library preperation | |
| | • Normalised GS-FLX | |
| | • Non-normalised GAII | |
| 2. | Raw sequence data (GS-FLX & GAII) | fasta & qual |
| 3. | Clustering of GS-FLX reads | fasta, qual |
| 4. | Mapping of GAII reads to the clusters | xls, text |
| 5. | Clustering table / assembly | xls, text |
| 5. | Statistics, summary | pdf, txt |

© 2007-2009

# Normalisation of cDNA

**cDNA Lib** › Raw Data › Clustering › GA Mapping › Visualization



**Tissue or Pool of Tissues**

mRNA
(active genes)

cDNA
synthesis

ds-cDNA
(amplified by
PCR)

denaturation

ss-cDNA

second
round of
normalization

reassoziation

ds-cDNA

removal of
ds-cDNA

ss-cDNA

Normalized cDNA

© 2007-2009

# Transcriptome: cDNA

cDNA Lib → Raw Data → Clustering → GA Mapping → Visualization

**Beispiele**: Chr2:10000..20000.

[Banner ausblenden] [Bookmark für diese Ansicht] [Link zur Abbildung dieser Ansicht] [Hilfe]

⊟ **Suche**

**Landmark oder Region**: Chr2:575001..625000

**Daten Quelle**: Illumina ChIP-Seq Database

**Dumps, Suchen und andere Operationen**: Annotieren Restriction Sites

**Scroll/Zoom**: Anzeigen 50 kbp ☐ **Umdrehen**

⊟ **Overview**

⊟ **Details**

## Mapping

© 2007-2009

# Acknowledgements

**Chris Bauser**

**Daniela Esser**

**Heike Hegele**

**Jiri Ködding**

**Lenia Beck**

**Martina Beck**

**Robert Greither**

**Thomas Pohl**

**Ulrike Schöck**

**Yadhu Kumar**

© 2007-2009

Thank you

www.gatc-biotech.com

© 2007-2009