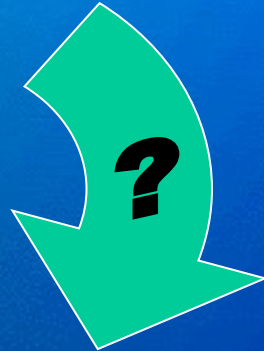
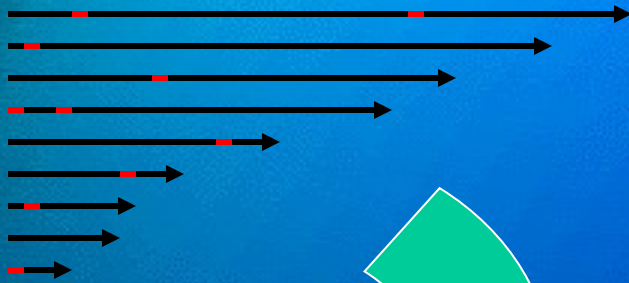


Assembling Sequences Using Trace Signals and Additional Sequence Information

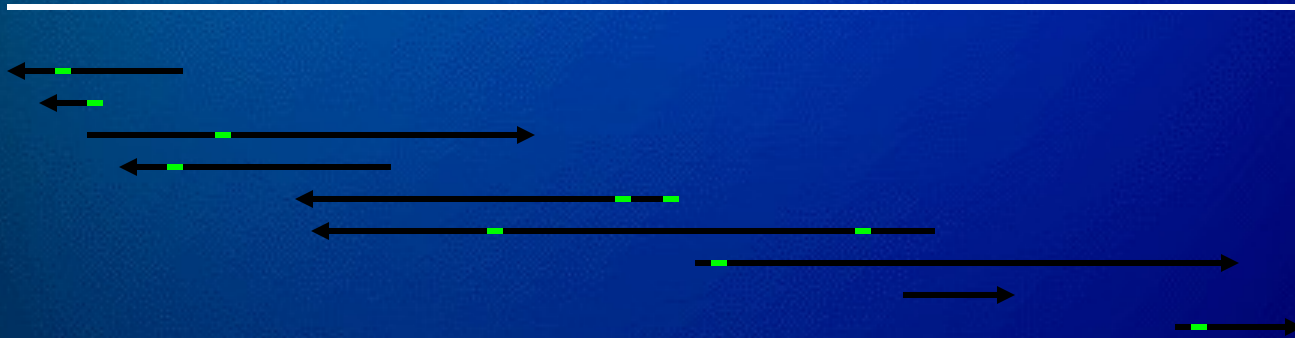
Bastien Chevreux, Thomas Pfisterer,
Thomas Wetter, Sandor Suhai

Deutsches Krebsforschungszentrum
Heidelberg

Problem definition



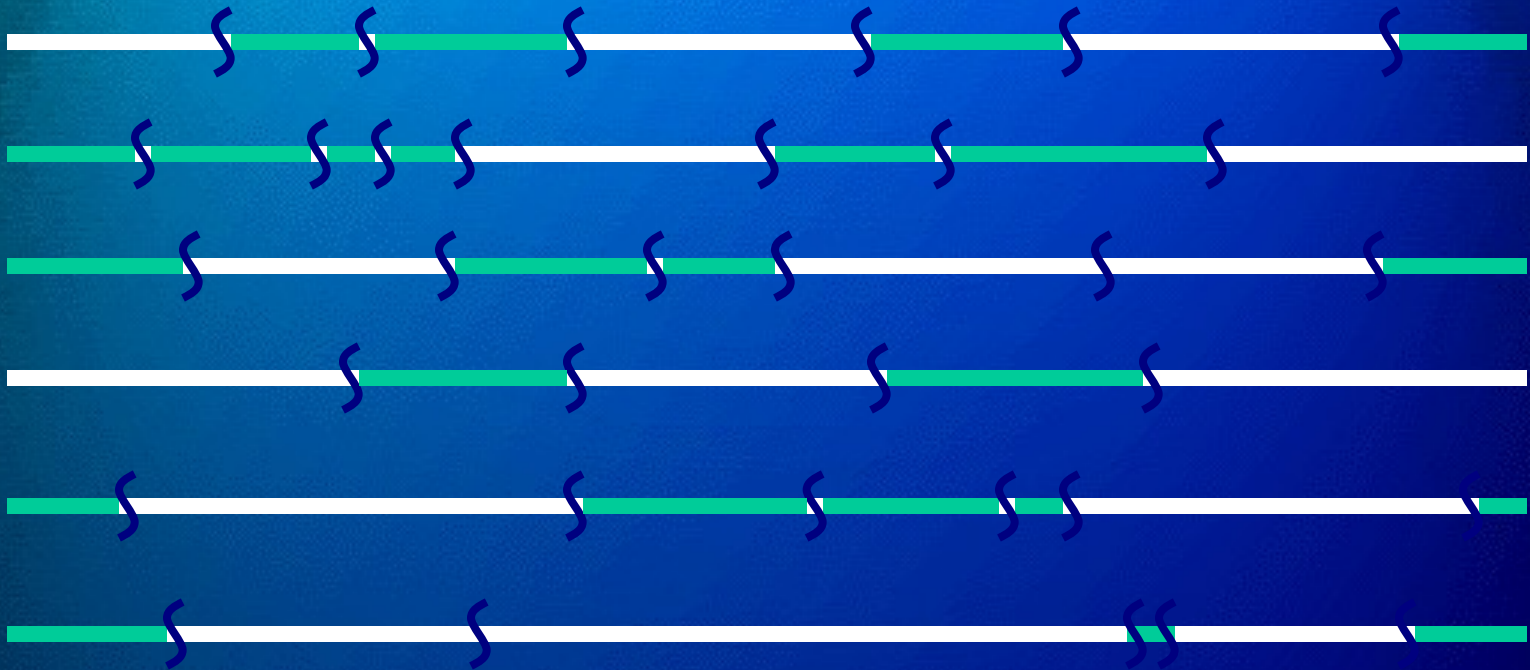
Assembly
&
Editing



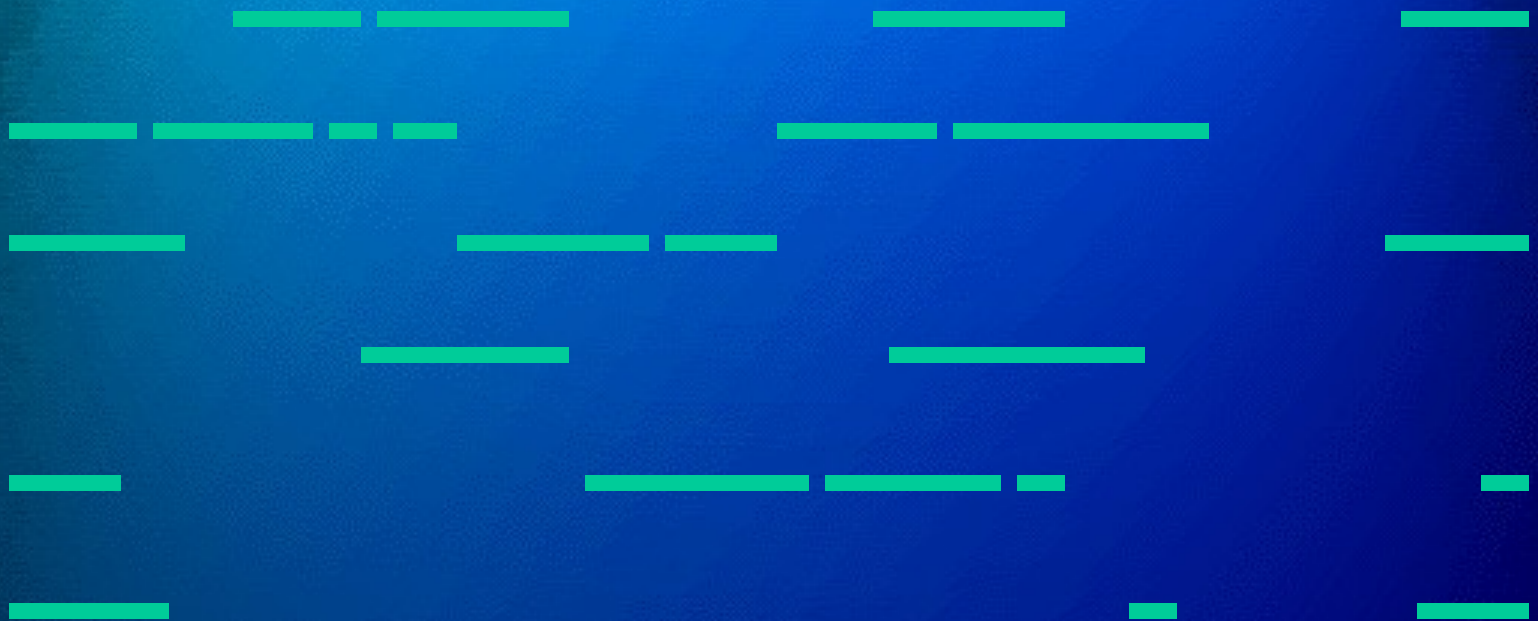
Introduction



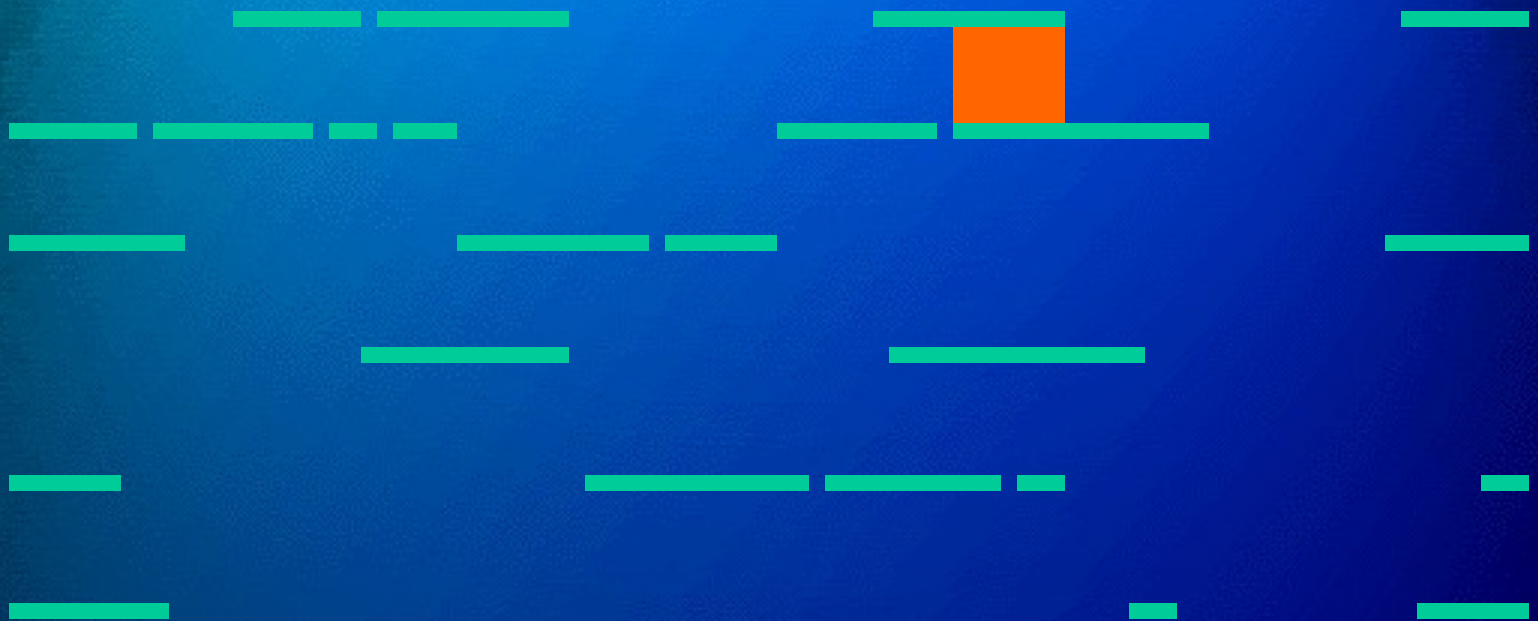
Introduction



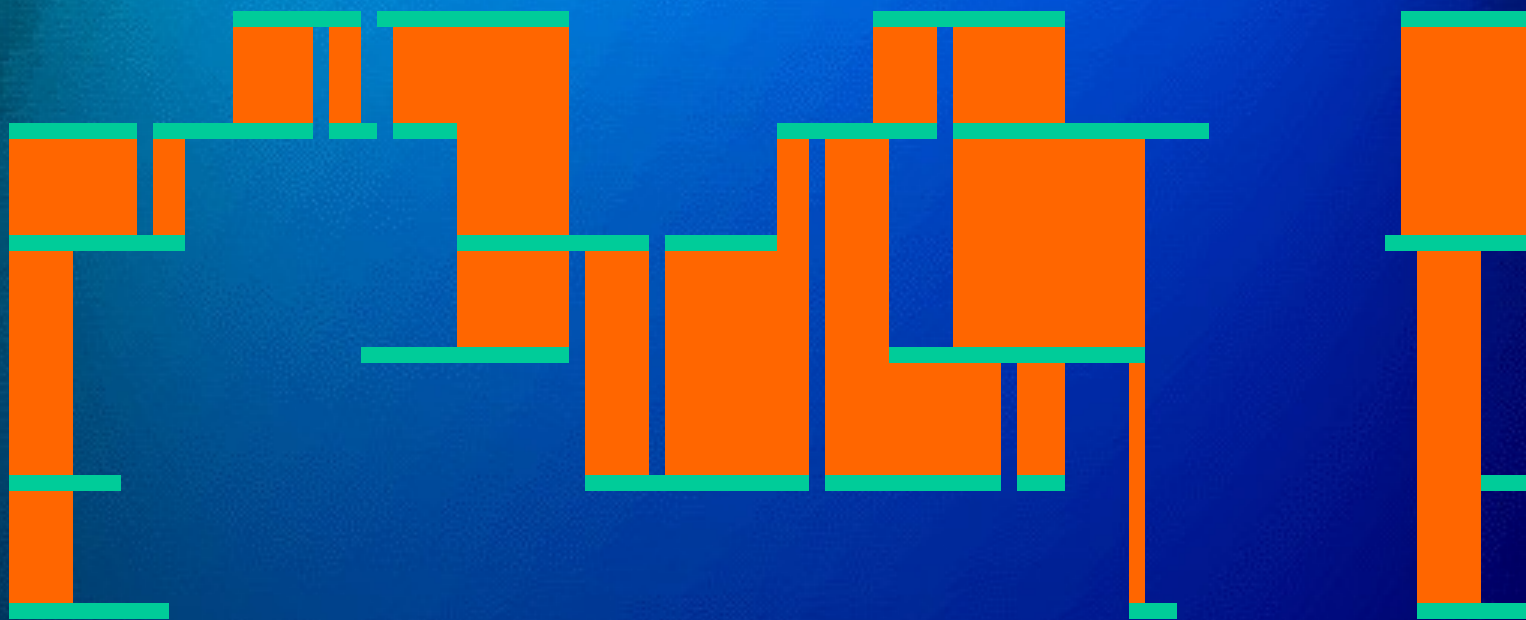
Introduction



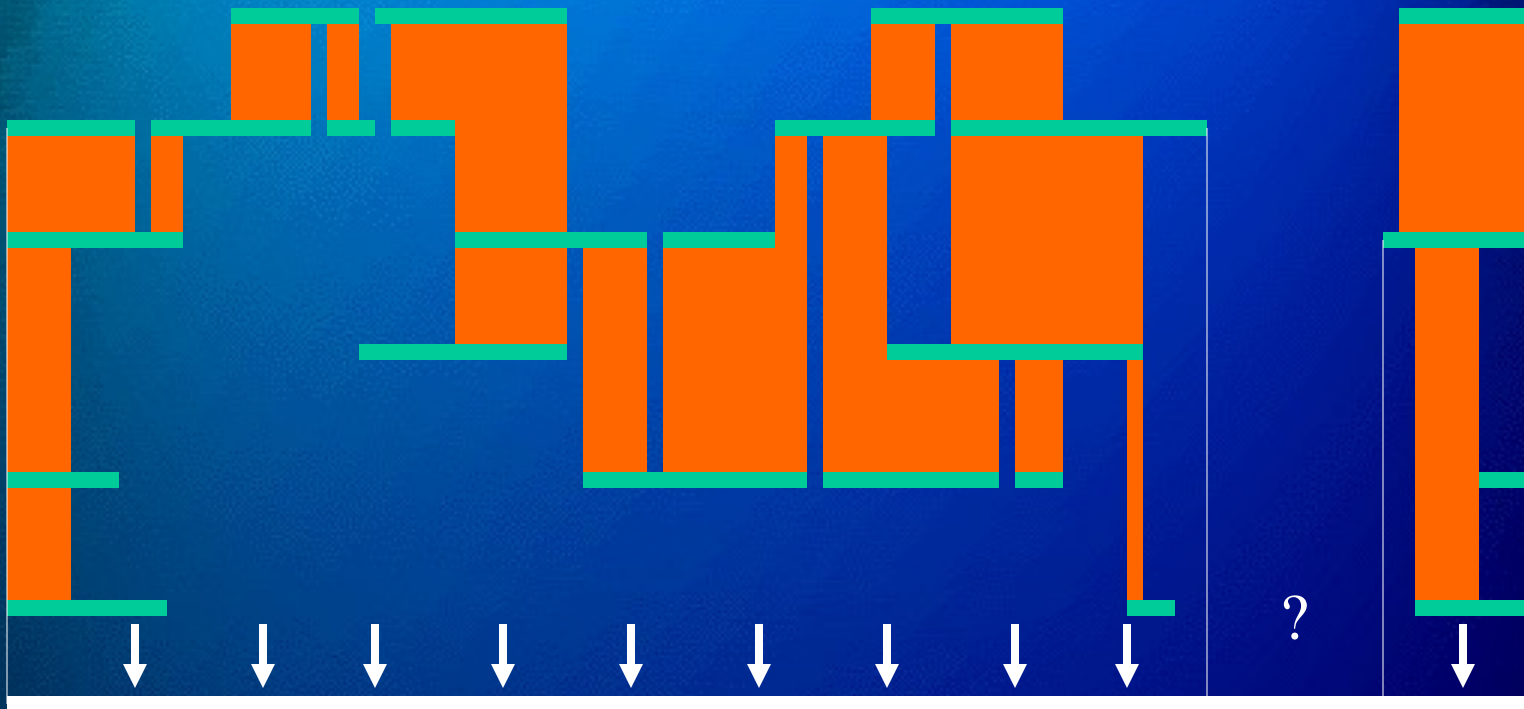
Introduction



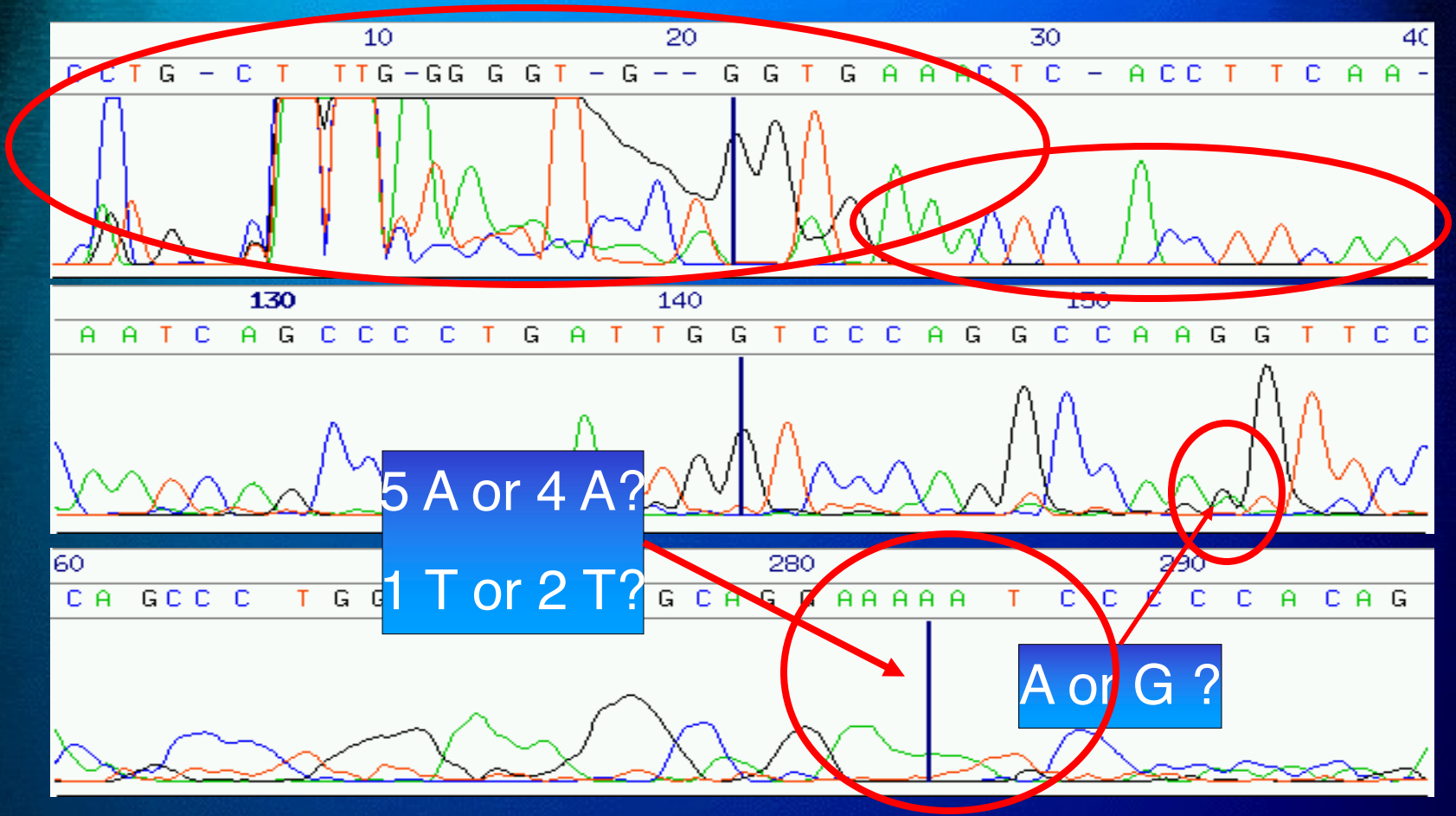
Introduction



Introduction



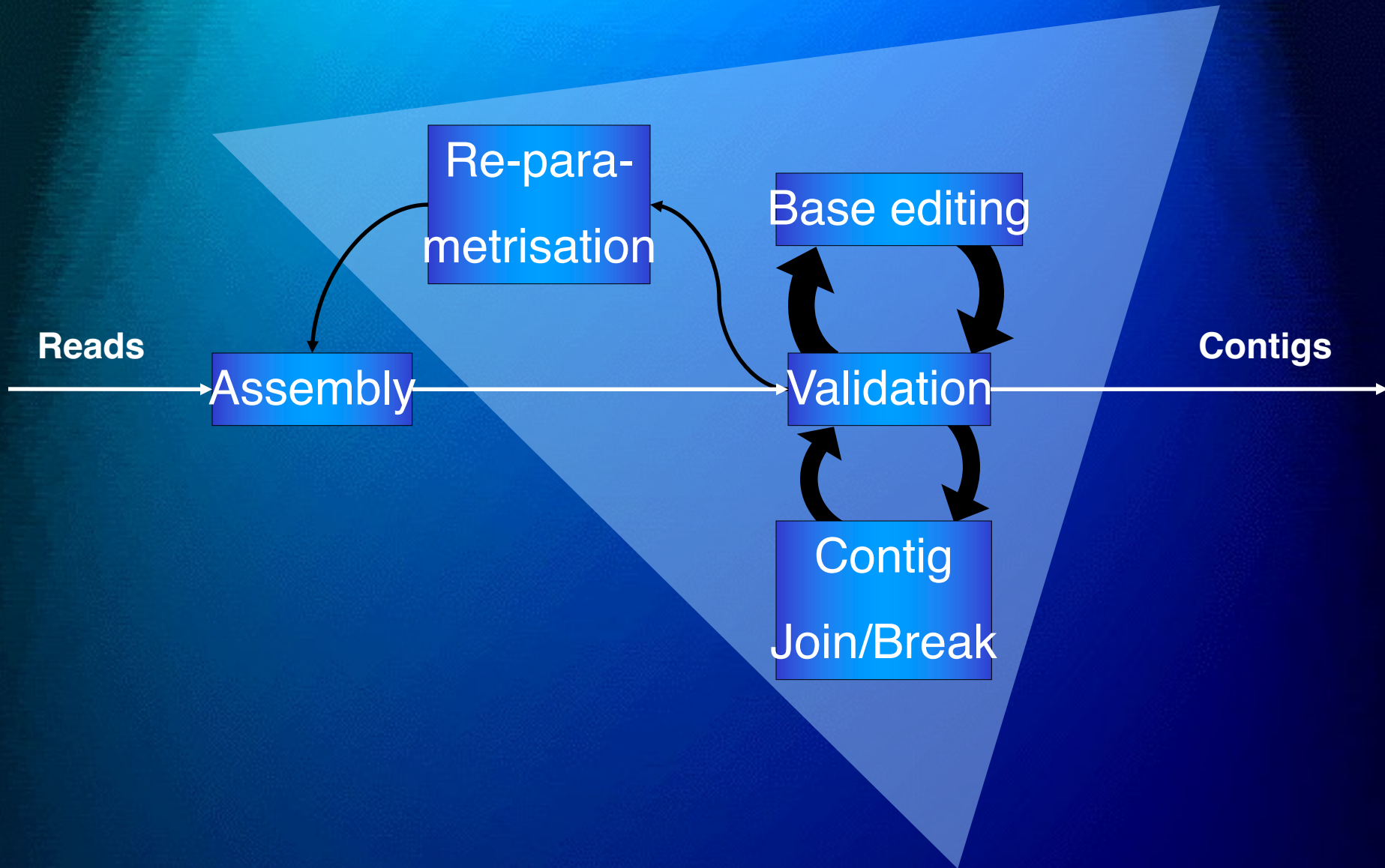
Signal problems



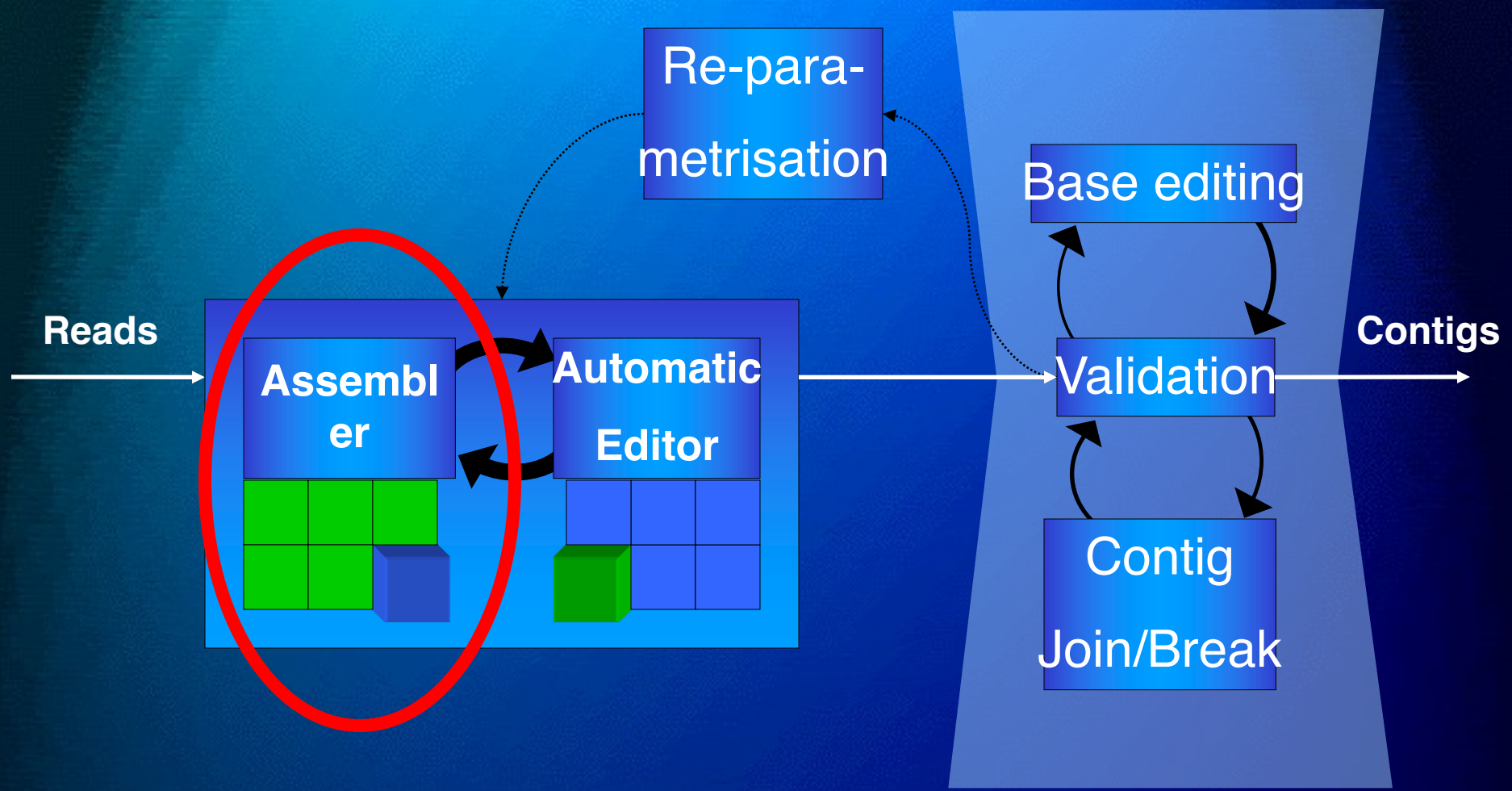
DNA problems

- Chemical properties
 - Coiling of DNA
 - Problems with dye chemistry
- Repetitive elements
 - Standard short term repeat (ALU, REPT etc.)
 - Long term repeats of sometimes several kb

Conventional assembly



Integrated Assembler-Editor



Assembler: Input

- Collection of reads
 - unknown relationship
 - unknown direction
- Each read
 - unknown error distribution
 - sequencing vector tagged
 - trace signal information
 - opt. base quality values
 - opt. quality clipping, marking HCRs (High Confidence Regions)
 - opt. standard repeats tagged
 - opt. template information

Assembly: Framework

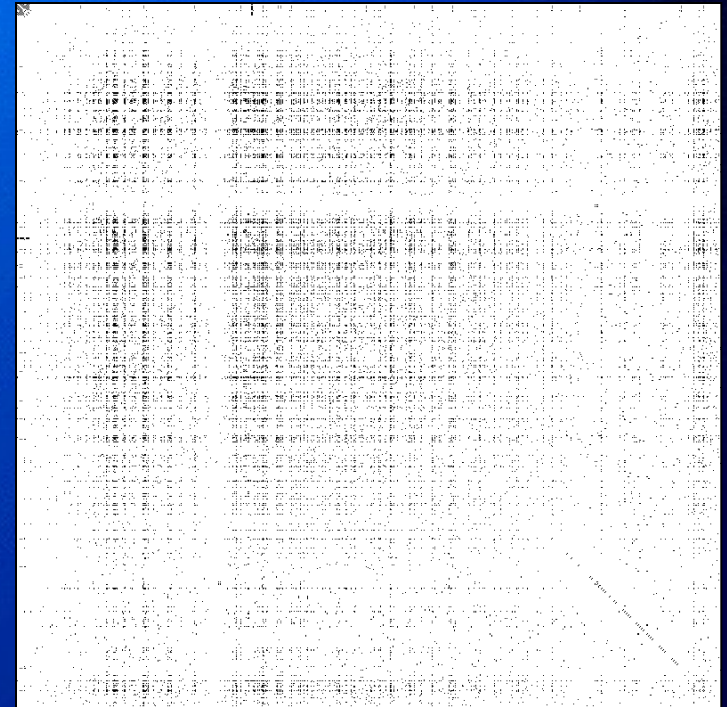
- Establishing relationships of each read against each other results in full oversight over the whole assembly
- Problem: k reads \rightarrow time complexity $O(k^2)$
- Fast read comparison routines needed
- Smith-Waterman has $O(mn)$, very slow

DNA-SAND algorithm

- Shift-AND algorithm: fault tolerant, $O(cmn)$
- modified Shift-AND for read comparison, DNA-SAND: fault tolerant, $O(cn)$ with $0 < c < 12$
- high sensitivity and specificity
 - less than 0.75% missed overlaps
 - around 45-50% false positive hits

Assembly: Framework

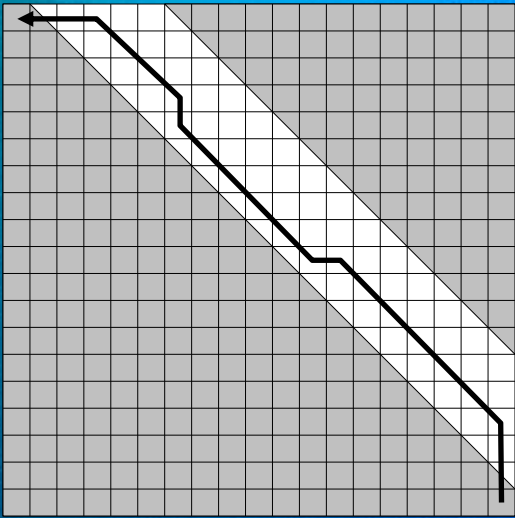
- Fault tolerant
- Sandsieve principle: obvious mismatches discarded, potential matches remembered
- Check each read in forward and reverse complement direction



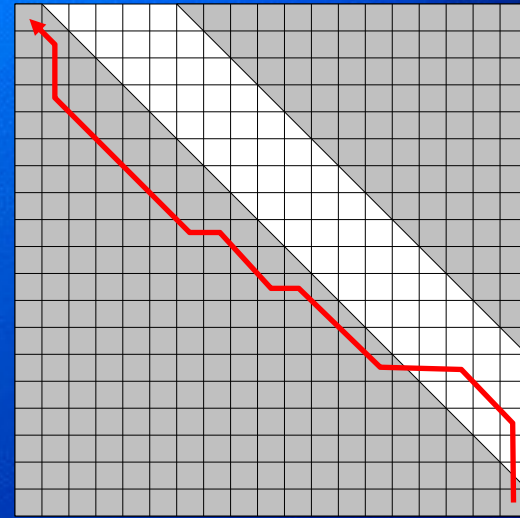
Overlap confirmation

- Evaluates potential overlaps
- Standard (banded) Smith-Waterman algorithm: $\max(O(bm), O(bn))$
- Rough calculation of SW match quality, eliminating false positive DNA-SAND matches
- Calculate an “alignment weight” for accepted overlaps

Overlap confirmation



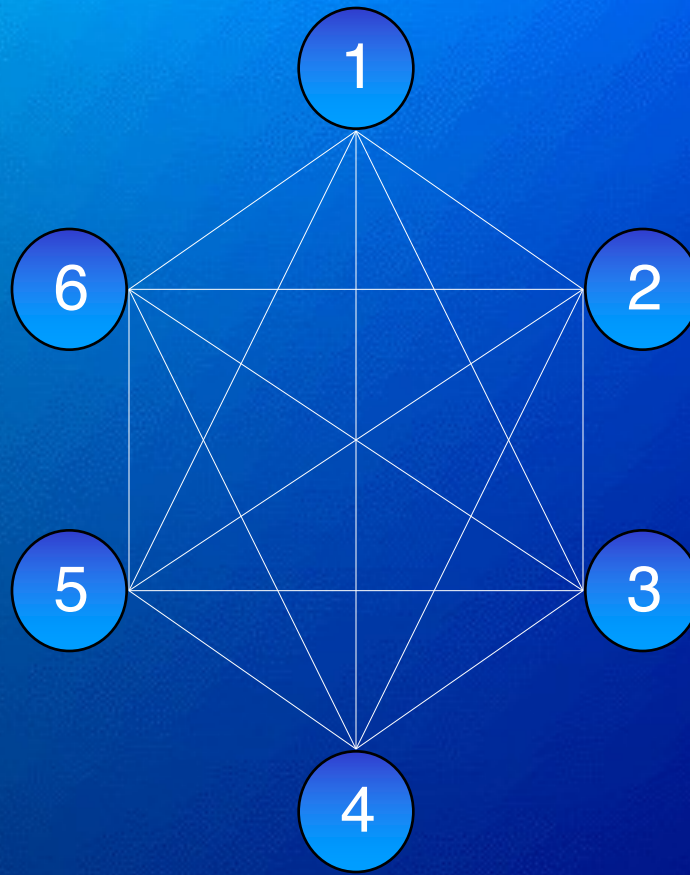
- Accepted match
 - Overlap: 196 bases
 - Score: 180
 - Score ratio: 92%
- Weight: 151817



- Rejected match
 - Out of band!
 - Overlap: 204 bases
 - Score: 133
 - Score ratio: 65%

Building a weighted graph

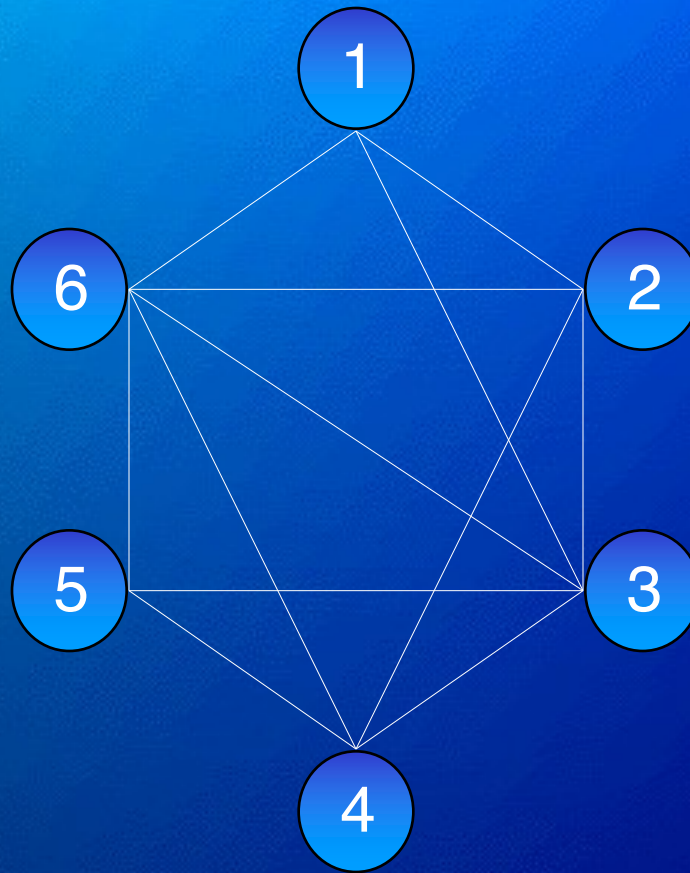
Example:
6 reads



All possible
overlaps for 2
reads

Building a weighted graph

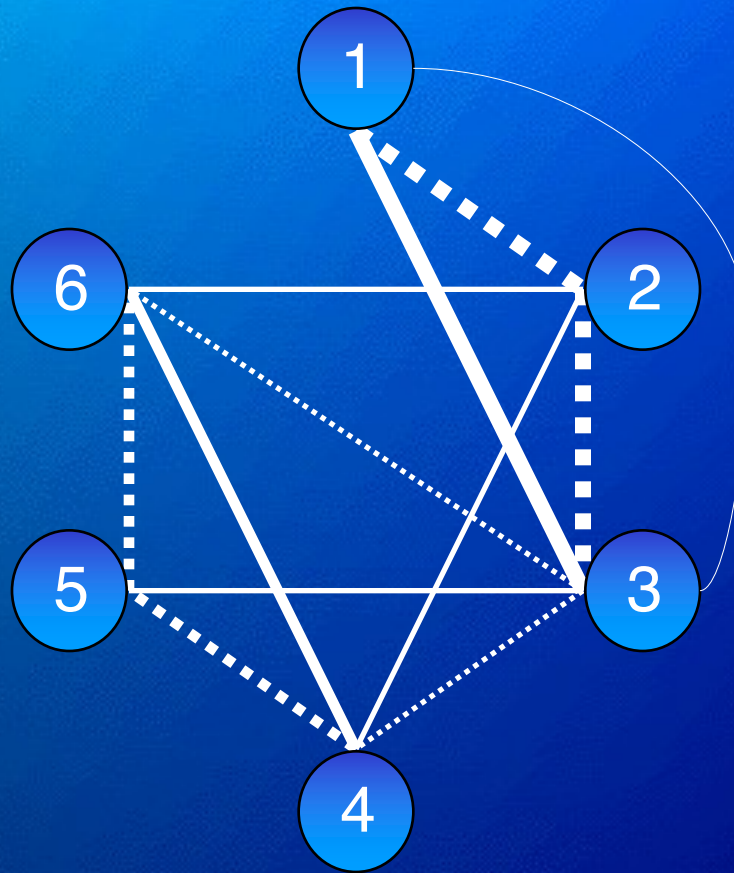
Pruned by
DNA-SAND



Building a weighted graph

Smith-Waterman

- Prune
- Attribute
 - direction
 - weight



Building contigs

- Multiple alignment is too slow
- Building a consensus by iteratively aligning reads against existing consensus
- Important:
 - Order of read alignments
 - Finding good alignment candidates
 - Possibility to reject candidates

Interaction: Pathfinder & Contig

- Pathfinder:
 - search good starting point for contig building
 - find good alignment candidates to add to existing contig
 - always inspect alternative paths in overlap graph
- Contig:
 - accept reads that match to existing consensus
 - reject reads that do not match
 - find inconsistencies that 'build up slowly' and mark these

Pathfinder: Strategy

- Finding starting points:
 - Search for node with a high number of reasonably weighted edges
 - Exclude edges below threshold
- Finding next alignment candidate:
 - Find reads with best nodes in contig
 - Recursively analyse best edges in graph

Contig: Strategy

- Align given read of given edge to existing contig at approximated position
- Accept read that match
- Reject reads that introduce
 - significantly higher error rates in contig than predicted by weighted edge
 - many non-editable errors in repetitive regions
 - inconsistencies with given template insert sizes

Contig: Raw

Contig Editor: -98 U13a04g12.t1

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

<< < > >>

	610	12620	12630	12640	12650	12660	12670	12680	12	
210 U13a02a07.t1	AACTTCAGGTTCAAA	*GACTTC	CC	*TGT	GAGCTTTTTCTTA	AATAAAGAAGAA	AATGTCAATATTACA			
211 U13a03f11.t1	AACTTCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTTTTTCTTA	AATAAAGAAGAA	AATGTCAATATTACA	AAGAA	*C		
-212 U13a01f04.t1	AACTTCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTTTTTCTTA	AATAAAGAAGAA	AATGTCAATATTACA	AAGAA	*C		
213 U13a06c11.t1	AACTTCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTTTTTCTTA	AATAAAGAAGAA	AATGTCAATATTACA	AAGAA	*C		
233 U13a09b03.t2	AACT									
-226 U13e07d01.t2	AACTTCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTTTTTCTT	AT	TAAAGAAGAA	AATGTCAATATTACA	AAGAA	CC	
-219 U13a05c12.t1	AACTT	AGGTTCA	AT	*GACTTCACC	*TGT	*GAGCTTTTTCTTA	AATAAAGAAGAA	AATGTCAATATTACA	AAGAA	*C
-214 U13a02a01.t1	AACTTCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTTTTTCTTA	AT	AAGAAGAA	AATGTCAATATTACA	AAGAA	*C	
-216 U13a06a03.t1		TCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTT	TTCTTAATATAAAGAAGAA	AATGTCAATATTACA	AAGAA	*C	
215 U13a08d12.p1		TTCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTTTTTCTTA	AATAAAGAAGAA	AATGTCAATATTACA	AAGAA	*C	
241 U13a08d12.t1		AAATCC	ACTTCACC	TGT	*AGCTTTT	CTTAATATAAAGAG	GAAAT	ATGTCAATATTACA	AAGAA	*C
-217 U13a01c10.t1							GAAATAATGTCAATATTACA	AAGAA	*C	
CONSENSUS	----	AACTTCAGGTTCAAA	*GACTTCACC	*TGT	*GAGCTTTTTCTTA	AATAAAGAAGAA	AATGTCAATATTACA	AAGAA	*C	

Reading:U13a05c12.t1(#219) Length:212(861) Vector:m13mp18.vec Clone:unknown Chemistry:primer Primer:unknown

Contig: Edited

Contig Editor: -98 U13a04g12.t1 <2>

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

<< < > >>

	30	12540	12550	12560	12570	12580	12590	12600	126
210 U13a02a07.t1	AACTTCAGGTTCAAA	*GACTTC	aCCTGT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACA	
211 U13a03f11.t1	AACTTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC	CT	
-212 U13a01f04.t1	AACTTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC	CT	
213 U13a06c11.t1	AACTTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC	CT	
233 U13a09b03.t2	AACT								
-226 U13e07d01.t2	AACTTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTTTT	CTT	a	a	TAAAGAAGAATAA	TGTCAATATT	ACAAGAACC
-219 U13a05c12.t1	AACTTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC	CT	
-214 U13a02a01.t1	AACTTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC	CT	
-216 U13a06a03.t1	tTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTT	tT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC	CT
215 U13a08d12.p1	TTCAGGTTCAAA	*GACTTCACCTGT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC	CT	
241 U13a08d12.t1	AAAT	CCACTTCACCT	GT	g	AGCTTTT	t	CCTAATATAA	AGAAGAATAA	TGTCAATATT
-217 U13a01c10.t1							a	GAAATAAT	TGTCAATATT
CONSENSUS	----	AACTTCAGGTTCAAA	*GACTTCACCT	GT	GAGCTTTTT	CCTAATATAA	AGAAGAATAA	TGTCAATATT	ACAAGAACC

Tag type:ED_C Direction:+ Comment:"Alter Base at 12660 1 # G a U13a08d12.t1 79"

Contig: Raw

Contig Editor: -416 U13a09h09.t1

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

<< < > >>

	640	650	660	670	680	690	700	710
406 122533c.t7	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCT							
405 122533b.t7	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCT							
-399 122533a.t3	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
401 230375b.t2	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
-400 230375a.t3	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
-402 230375c.t3	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
398 230375b.t1	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
-403 122533c.t3	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
-404 230375b.t3	AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
353 U13a03g02.t1	AAAGTACTATAAGAATAATTTATAGCAATGAGTTTAAAACTCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
-352 U13e07d02.t2	AGAATAATTTATGCAATGGTTTAAAACTCAA-TTTTTATTGC-TTCTCACCAGCTGC-AAGTGTTT							
-415 122533b.t3	AGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
-357 U13a04c06.t1	AATAATTTATAGCAATGAGTTTAAAACTCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							
CONSENSUS	----AAAGTACTGTAAGAATAATTTATAGTAATGAGTTTAAAAATCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT							

Contig:U13a09h09.t1(#416) Length:22938

Contig: Edited

Contig Editor: -416 U13a09h09.t1 <2>

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

<< < > >>

	640	650	660	670	680	690	700	710
406 122533c.t7	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCT					
405 122533b.t7	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCT					
-399 122533a.t3	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
401 230375b.t2	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
-400 230375a.t3	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
-402 230375c.t3	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
398 230375b.t1	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
-403 122533c.t3	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
-404 230375b.t3	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
353 U13a03g02.t1	AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
-352 U13e07d02.t2	AGAATAATTTAT	CAATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
-415 122533b.t3	AGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
-357 U13a04c06.t1	AATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					
CONSENSUS	----AAAGTACTGTAAGAATAATTTATAG	AATGAGTTTAAAA	CAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTT					

Reading:U13a04c06.t1(#357) Length:453(915) Vector:m13mp18.vec Clone:unknown Chemistry:primer Primer:unknown

Contig Editor: -435 U13a04d08.t1

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

<< < > >>

	2310	2320	2330	2340	2350	2360	2370	2380
406 U13a06d03.t1	AAAGTACTAT	AGAATAATTTATAG	CAATGAGTTTAAACTCAACTTTTTATTGCCTTC	*TCACCAGCTGCAAAGTGTT				
407 U13e07f02.t2	AAAGTACTATAAGAATAATTTATAG	*CAATGAGTTTAAACTC	AACTTTTTATTGCCTTC	*TCACCAGCTGCAAAGTGTT				
411 U13a05b10.t1	AAAGTACTATAAGAATAATTTATAG	GCAAGGAGTTTAAACTCAACTTTTTATTGCCTTC	*TCACC	GCTGCAAAGTGTT				
405 U13a08f11.t1	AAAGTACTA	AGAATAATTTATA*GCAATGAGTT	AAACTCAACTTTTTATTGCCTTC	*TCACCA	CTGCAAAGTGTT			
404 U13a09h09.t1	AAAGTACTATAAGAATAATTTATA*GCAATGAGTTTAAACT	AACTTTTTATTGCCTTC	ATCAC	AGCTGCAAAGTGTT				
331 U13a03g02.t1	AAAGTACTATAAGAATAATTTATA*G	CAATGAGTTTAAACTCAACTTTTTATTGCCTTC	*TCACCAGCTGCAAAGTGTT					
-330 U13e07d02.t2		AGAATAATTTAT	*GCAATGAGTTTAAACTCAAC	TTTTTATTGCCTTC	*TCACCAGCTGC	AAGTGTT		
-335 U13a04c06.t1		AATAATTTATA*G	CAATGAGTTTAAACTCAACTTTTTATTGCCTTC	*TCACCAGCTGCAAAGTGTT				
CONSENSUS	----	AAAGTACTATAAGAATAATTTATA*GCAATGAGTTTAAACTCAACTTTTTATTGCCTTC	*TCACCAGCTGCAAAGTGTT					

Tag type:PRMB Direction:+ Comment:"Possible Repeat Marker Base found by MIRA."

Contig Editor: -435 U13a04d08.t1

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

<< < > >>

	2300	2310	2320	2330	2340	2350	2360	2370
406 U13a06d03.t1	AAAGTACTAT	aAGAATAATTTATAGCAATGAGTTTAAACTCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTT						
407 U13e07f02.t2	AAAGTACTATAAGAATAATTTATAGCAATGAGTTTAAACTC	AACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTT						
411 U13a05b10.t1	AAAGTACTATAAGAATAATTTATAGCAATGAGTTTAAACTCAACTTTTTATTGCCTTCTCACC	aGCTGCAAAGTGTT						
405 U13a08f11.t1	AAAGTACTA	aAGAATAATTTATAGCAATGAGTT	AAACTCAACTTTTTATTGCCTTCTCACC	aGCTGCAAAGTGTTTT				
404 U13a09h09.t1	AAAGTACTATAAGAATAATTTATAGCAATGAGTTTAAACT	AACTTTTTATTGCCTTCTCAC	aAGCTGCAAAGTGTTTT					
331 U13a03g02.t1	AAAGTACTATAAGAATAATTTATAG	CAATGAGTTTAAACTCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTT						
-330 U13e07d02.t2		AGAATAATTTAT	aGCAATGAGTTTAAACTCAAC	TTTTTATTGCCTTCTCACCAGCTGC	AAGTGTTTT			
-335 U13a04c06.t1		AATAATTTATAG	CAATGAGTTTAAACTCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTT					
CONSENSUS	----	AAAGTACTATAAGAATAATTTATAGCAATGAGTTTAAACTCAACTTTTTATTGCCTTCTCACCAGCTGCAAAGTGTTTT						

Tag type:PRMB Direction:- Comment:"Possible Repeat Marker Base found by MIRA."

Contig Editor: -98 U13a04g12.t1

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

<< < > >>

	1740	1750	1760	1770	1780	1790	1800	1810
-80 U13a06b01.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
81 U13a02e11.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
104 U13a06h05.t1	AGACACCA							
-79 U13a01e08.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
111 U13a02c01.t1	AGACACCAAAATAGTGGAT							
78 U13a07b08.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
-90 U13a06d12.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
-101 U13a08b12.t1								CTCCTG
CONSENSUS	----	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG						

Tag type:ED_C Direction:+ Comment:"Alter Base at 1747 1 # N g U13a06h05.t1 287"

Contig Editor: -98 U13a04g12.t1

< C: 1 > < Q: -1 > Insert Edit Modes Cutoffs Undo Next Search Commands Settings Quit Help

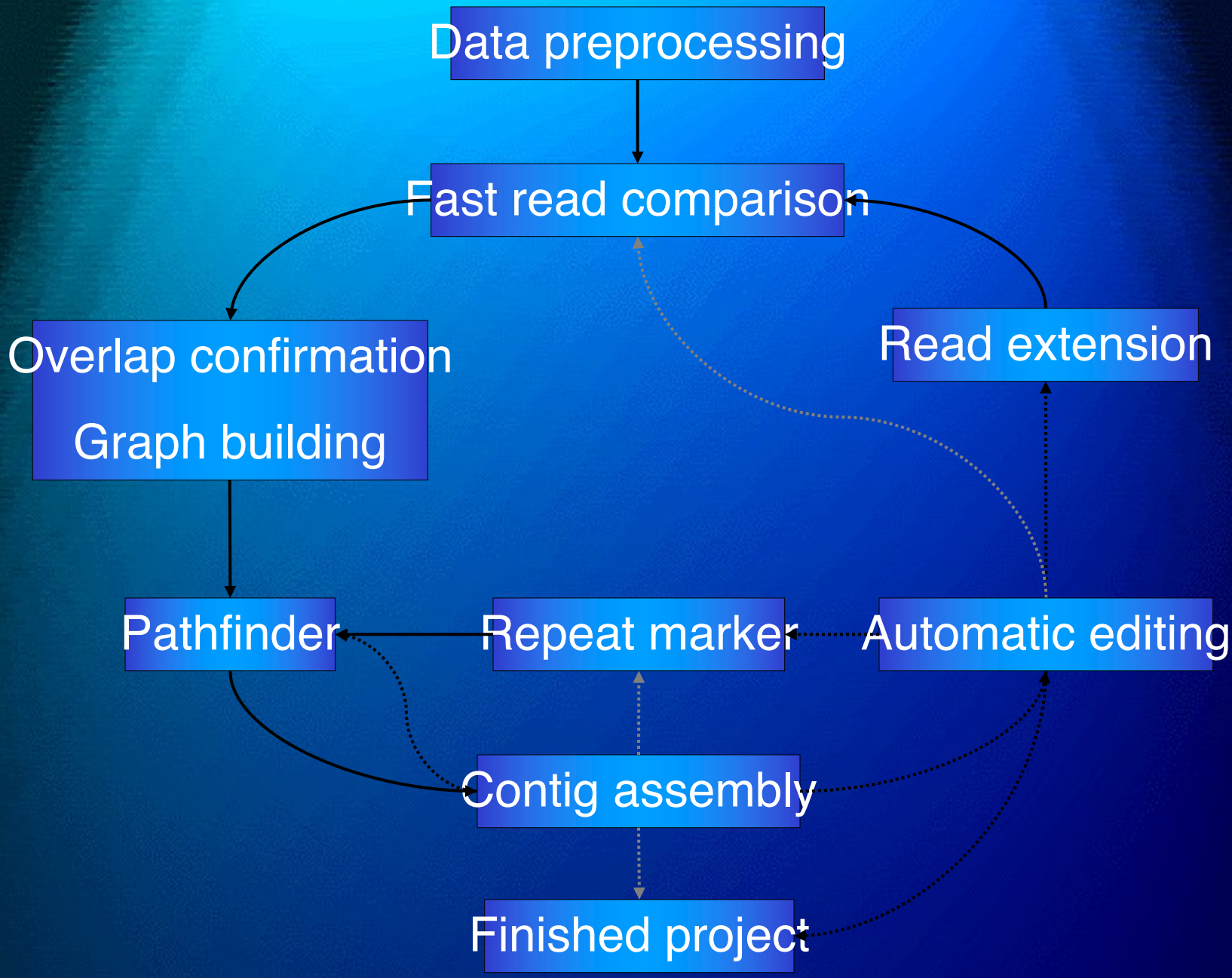
<< < > >>

	1740	1750	1760	1770	1780	1790	1800	1810
-80 U13a06b01.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
81 U13a02e11.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
104 U13a06h05.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
-79 U13a01e08.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
111 U13a02c01.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
78 U13a07b08.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
-90 U13a06d12.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
-101 U13a08b12.t1	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
-107 U13a01a07.t1	CAGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG							
-75 U13a04f02.t1								CAC
-131 U13a010c08.t1	GAATAA-GGTGGGAT-GGTAAAAA-AG---GCAA-TGGGGG-TAAGGGGGGAAA-AA--ATAGAA-CAAAGG-AAAAGAA							
-125 U13a08b01.t1	-T-TTCCCGT-AAGAT--GA-G-TTCGAGATGA-CA-CCTCTT-CACA-CAA-GAAGGG-ATA-A-G-C--AAGG-AAG							
CONSENSUS	----	AGACACCAAAATAGTGGATGCCCAAGTCCCTGATAGAAAAAATTGCATAGTATCTGCATATGACCTATGCAATCCTCCTG						

Tag type:ED_C Direction:+ Comment:"Alter Base at 1747 1 # N g U13a06h05.t1 287"

Extending HCRs

- 'beef up' existing contigs; trivial, very fast
- extend existing contigs; simple, quick
- find new contigs to build; bold, slow



Status

- beta-testing almost completed
- assembler & editor in use to assemble projects up to 10.000 reads
- first evaluation: human finished 35kb project (Golden Standard)
without fine-tuning assembled contigs have 99,9x% identity
- whole genome shotgun with 23.000 reads in preparation
- other applications like EST clustering?

Canonical Homepage

<http://www.dkfz-heidelberg.de/mbp-ased/>

Acknowledgements

Prof. Rosenthal, Matthias Platzer, Uwe Menzel and the
IMB Jena genome sequencing centre

Bernd Drescher and Lion Biosciences AG, Heidelberg