Structural Variation Discovery
‘Next-Generation Tools’ for Next-Generation Data

Tobias Marschall
CWI, Amsterdam

NBIC NGS Meeting
June 14, 2013
Paired-End Sequencing

1. Generate sequencing library

2. Sequence fragments starting from both ends
Re-Sequencing given a Reference Genome

- **Top:** reference genome
- **Bottom:** aligned read pairs
- **In red:** differences between read and reference
Genetic Variations

**Single nucleotide polymorphisms (SNPs)**

- CCCAGCACTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT
- CCCAGCACTTTGGGAGGTCAAGGTGGGGGGAGGAAATAGCTTAAGCCCAGGAGT

**Deletion**

- CCCAGCACTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT
- CCCAGCACTTTGGGAGGTCAAGGTGGGGGGAGGAAATAGCTTAAGCCCAGGAGT

**Insertion**

- CCCAGCACTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT
- CCCAGCACTTTGGGAGGTCAAGGTGGGGGGAGGAAATAGCTTAAGCCCAGGAGT

**Translocation**

- CCCAGCACTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT
- CCCAGCACTTTGGGAGGTCAAGGTGGGGGGAGGAAATAGCTTAAGCCCAGGAGT

**Further variations:** inversions, duplications, ...
# Approaches for Structural Variation Discovery

## Possible approaches / information sources
- Coverage based
- Internal segment length based
- Split read mapping
- (Local) assembly

## Current challenges
- Small/mid-size deletions
- Repetitive regions
## Research Agenda

### Goals
- Push individual approaches to their limits
- Combine them
- Be trio/population aware

### Approaches
- **CLEVER**: Internal segment based (Bioinformatics, 2012)
- **LASER**: Split read mapping (arXiv preprint 1303.3520)
- **MATE-CLEVER**: Combining both, adding trio-awareness and genotyping (HitSeq/Bioinformatics, 2013), used for GoNL

### Next year: (local) assembly
Clever: Clique-Enumerating Variant Finder

Reference
Marschall, Costa, Canzar, Bauer, Klau, Schliep und Schönhuth
Software (GPL): http://clever-sv.googlecode.com
Discovering Deletions

Reference genome

 CCCAGCACTTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

Deletion

⇒ observing a longer internal segment

Insertions

⇒ observing a shorter internal segment

Null model:

internal segment length: normal distribution $N_{\mu, \sigma}$
Discovering Deletions

Reference genome
CCCAGCAGTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

Sequenced genome
CCCAGCAGTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

Deletion = observing a longer internal segment
Insertions = observing a shorter internal segment

Null model: internal segment length: normal distribution $N_{\mu,\sigma}$
Discovering Deletions

Reference genome
CCCAGCACCTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

Sequenced genome
CCCAGCACCTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

Deletion = observing a longer internal segment
Insertions = observing a shorter internal segment

Null model: internal segment length: normal distribution $\mathcal{N}(\mu, \sigma)$
Discovering Deletions

**Reference genome**

CCCAGCACTTTTGGGAGGCCAAGGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

**Sequenced genome**

CCCAGCACTTTTGGGAGGCCAAAAATTGCTTAAGCCCAGGAGT

Deletion $\Rightarrow$ observing a longer internal segment

Insertions $\Rightarrow$ observing a shorter internal segment

Null model: internal segment length: normal distribution $\mathcal{N} \mu, \sigma$
Discovering Deletions

**Reference genome**
CCCAGCAGTTTTGGGAGGCAAGGTTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

**Sequenced genome**
CCCAGCAGTTTTGGGAGGCAAGGTTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT

**Deletion** ⇒ observing a longer internal segment

**Insertions** ⇒ observing a shorter internal segment
Discovering Deletions

**Reference genome**

```
CCCAGCACTTTTGGGAGGCCAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT
```

**Sequenced genome**

```
CCCAGCACTTTTGGGAGGCCAAAGGTGGGGGGAGGAAATTGCTTAAGCCCAGGAGT
```

**Deletion** ➞ observing a longer internal segment

**Insertions** ➞ observing a shorter internal segment

**Null model:** internal segment length: normal distribution $\mathcal{N}_\mu,\sigma$
CLEVER: Commonly Discordant Alignments

IGV Screenshot: Deletion

**Red reads:** Insert $\geq \mu + 2.5\sigma$
A Less Clear Situation

IGV Screenshot: Deletion

**Red reads:** Insert $\geq \mu + 2.5\sigma$
Prior Approaches (Internal Segment Length Based)

Problem 1
Reads with *concordant* alignments are discarded
⇒ Cannot detect small events

Problem 2
*Multiply mapped* reads are not handled properly or discarded altogether
⇒ Ignores repetitive regions
**Approach**

Nodes: All(!) alignments

Edges: Alignments supporting **the same allele**

Find all **maximal cliques** (max-cliques) and make predictions
Edges

Incompatible Alignments (NO edge):

1. Too large length difference
   \[ I(A) \approx \mu \]
   \[ I(B) > \mu \]

2. Long internal segments, but small overlap
   \[ I(A) > \mu \]

Compatible Alignments (edge):

3. Average internal segments lengths, small overlap
   \[ I(A) \approx \mu \]
   \[ I(B) \approx \mu \]
   \[ O(A,B) \]

4. Long internal segments, sufficient overlap
   \[ I(A) > \mu \]
   \[ I(B) > \mu \]
   \[ O(A,B) \]
   \[ I(A,B) > \mu \]
Edge Criteria (Formal)

Let $X$ be $\mathcal{N}(0, 1)$-distributed.

- Edges: Two-sided two sample Z-test for statistically compatible insert size

$$P\left( X \geq \frac{\Delta_{12}}{\sqrt{2}\sigma} \right) \leq \alpha = 0.1$$

and one-sided one sample Z-test for statistically consistent overlap

$$P\left( X \geq \frac{\sqrt{2}(U_{12} - \mu)}{\sigma} \right) \leq \alpha = 0.1$$

- Edges reflect that one cannot reject the hypothesis that the reads behind the alignments stem from the same allele.
Fast Implementation

Techniques
- Active alignments: binary search tree (sorted by insert length)
- Cliques: store as *bit-vectors* over active alignments
  - Clique intersection *bit-parallel*
- Reorganize storage now and then

Runtime
- $30 \times$ coverage, *all* reads, up to $\approx 650$ alignments per read:
  - Around **20 minutes** for whole chromosome 1
Evaluation of Max-Cliques and Predictions

CLEVER

- **Multiply mapped reads**: use alignment score-driven probability distributions of reads over their alignments.
- **Each max-clique** $C$ (accounting for multiply mapped reads):

  \[
P_{H_0}(C) = \sum_{C \subset C} P(C\text{ correct and } C \setminus C\text{ incorrect}) \cdot P_{H_0}(C)
  \]

  where $H_0$ null hypothesis of no variation,

  \[
P_{H_0}(C) = P(X_{\mathcal{N}(0,1)} \leq \sqrt{|C|} \cdot \frac{\bar{I}(C) - \mu}{\sigma})
  \]

  reflects Z-test for sample of size $|C|$ for $C$ being commonly discordant.
- **After correction for multiple hypothesis testing**: predict indels from all significant cliques $C$. 
### CLEVER Results: Venter’s Genome

#### Deletions: Fixed-Threshold Statistics

<table>
<thead>
<tr>
<th>Length range</th>
<th>20–49</th>
<th>50–99</th>
<th>100-50K</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Algorithm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLEVER</td>
<td>93.2 / 74.2 / 6.4</td>
<td>84.9 / 82.4</td>
<td>86.0 / 68.2</td>
</tr>
<tr>
<td>BreakDancer</td>
<td>91.0 / 8.6 / 0.0</td>
<td><strong>94.6</strong> / 52.6</td>
<td>65.2 / 56.8</td>
</tr>
<tr>
<td>GASV</td>
<td>10.3 / 45.8 / 1.3</td>
<td>51.5 / 39.3</td>
<td>0.8 / 47.9</td>
</tr>
<tr>
<td>HYDRA</td>
<td>– / 0.1 / 0.0</td>
<td>– / 5.1</td>
<td>70.4 / 55.7</td>
</tr>
<tr>
<td>VariationHunter</td>
<td>82.9 / 11.3 / 0.3</td>
<td>76.4 / 77.4</td>
<td>55.9 / 63.2</td>
</tr>
<tr>
<td>PINDEL</td>
<td>68.4 / 73.9 / 9.7</td>
<td>76.5 / 40.0</td>
<td>84.1 / 39.4</td>
</tr>
<tr>
<td>SV-seq2</td>
<td><strong>100.0</strong> / 1.3 / 0.0</td>
<td>88.2 / 20.9</td>
<td>79.3 / 36.7</td>
</tr>
</tbody>
</table>
LASER: Sensitive Long-Indel-Aware Alignment of Sequencing Reads

Reference
Split-Read Alignments

Goal

Reliably detect

- Split alignments (i.e. covering a breakpoint)
- All other (regular) alignments
Wrong Alignments are the Root of (all) Evil

Red(dish): Misplaced Alignments
Overview

Observations

- Most read mappers geared towards speed, rather than alignment quality
- Need all possible alignments of (important) reads to draw conclusions
- Alignment scores are crucial to pick the right alignment

Strategy

- Use existing methods to find seed alignments
- Sensitive, base-quality-aware alignments
- Recalibrate alignment scores after we have seen all reads and know where SNPs and indels are
LASER – Step 1

Preprocessing: Extract prefix and suffix

Read 1: 100bp 35bp 35bp 1L 1R 2L 2R

Read 2: 100bp 35bp 2L 2R

Step 1: Map fragments to reference genome

Left forward anchor  Right forward anchor  Right backward anchor
LASER – Step 2 and Step 3

**Step 2:** Determine regions to search for more anchors

**Step 3:** Search for new anchors, discard duplicates

Iterate: Repeat these steps three times
LASER – Step 4 and Step 5

Step 4: Banded sequence alignment to extend anchors, discard duplicates

Step 5: Combine anchor alignments
## Probabilistic Alignment Scores

- **Compute** *probability distribution* over all possible alignment pairs

### Mismatches
- Mismatch cost based on base qualities (reported by sequencer)

### Indels
- Use empiric indel length distribution (on uniquely mapping reads)

### Insert size
- Use empiric insert size distribution (on uniquely mapping reads)
Alignment Score Recalibration

Reference genome

Read 1

Mismatch

Deletion
Alignment Score Recalibration

Reference genome

Read 1  
Mismatch

Read 2  
Deletion

Read 3
Read 4
Read 5
Alignment Score Recalibration

Reference genome

Read 1
Mismatch

Read 2
Deletion

Read 3
Read 4
Read 5

Solution: score recalibration

1. Align all reads independently, report multiple alignments
2. Call SNPs (and indels), considering all alignments
3. Change cost of these indels in alignment scores
4. Pick best alignment based on updated scores
LASER: Results

Venter’s genome / simulated reads
How many reads are mapped correctly?

Read categories
- Regular (no indels)
- **Short deletion** / insertion (1-20 bp)
- **Midsize deletion** / insertion (21-50 bp)
- **Long deletion** (50+ bp)
- Others (multiple indels, ending in insertion, etc.)
Percent correct reads with short deletions (1-20)

<table>
<thead>
<tr>
<th></th>
<th>LASER</th>
<th>BWA</th>
<th>Bowtie2</th>
<th>Stampy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary alignment only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully correct</td>
<td>69.4</td>
<td>67.5</td>
<td><strong>71.4</strong></td>
<td>52.8</td>
</tr>
<tr>
<td>Correct end points</td>
<td>5.7</td>
<td>2.4</td>
<td>1.7</td>
<td>20.7</td>
</tr>
<tr>
<td>One correct end point</td>
<td>22.0</td>
<td>28.5</td>
<td>25.8</td>
<td>25.6</td>
</tr>
<tr>
<td>Wrong but near</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Wrong</td>
<td>0.5</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Most favorable alignment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully correct</td>
<td><strong>75.8</strong></td>
<td>68.4</td>
<td>72.0</td>
<td>52.8</td>
</tr>
<tr>
<td>Correct end points</td>
<td>7.5</td>
<td>2.5</td>
<td>1.7</td>
<td>20.7</td>
</tr>
<tr>
<td>One correct end point</td>
<td>14.4</td>
<td>27.6</td>
<td>25.9</td>
<td>25.6</td>
</tr>
<tr>
<td>Wrong but near</td>
<td>0.0</td>
<td>0.4</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Wrong</td>
<td>0.1</td>
<td>0.7</td>
<td>0.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Percent correct reads with midsize deletions (21-50)

<table>
<thead>
<tr>
<th></th>
<th>LASER</th>
<th>BWA</th>
<th>Bowtie2</th>
<th>Stampy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary alignment only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully correct</td>
<td>38.0</td>
<td>11.1</td>
<td>0.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Correct end points</td>
<td>2.6</td>
<td>0.4</td>
<td>0.0</td>
<td>16.3</td>
</tr>
<tr>
<td>One correct end point</td>
<td>47.3</td>
<td>78.1</td>
<td>80.3</td>
<td>77.0</td>
</tr>
<tr>
<td>Wrong but near</td>
<td>0.9</td>
<td>4.6</td>
<td>2.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Wrong</td>
<td>1.0</td>
<td>1.8</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td><strong>Most favorable alignment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully correct</td>
<td>66.5</td>
<td>11.1</td>
<td>0.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Correct end points</td>
<td>9.3</td>
<td>0.4</td>
<td>0.0</td>
<td>16.3</td>
</tr>
<tr>
<td>One correct end point</td>
<td>13.0</td>
<td>79.5</td>
<td>82.5</td>
<td>77.0</td>
</tr>
<tr>
<td>Wrong but near</td>
<td>0.1</td>
<td>3.2</td>
<td>0.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Wrong</td>
<td>1.0</td>
<td>1.8</td>
<td>2.0</td>
<td>2.3</td>
</tr>
</tbody>
</table>
Percent correct reads with long deletions (50+)

<table>
<thead>
<tr>
<th></th>
<th>LASER</th>
<th>BWA</th>
<th>Bowtie2</th>
<th>Stampy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary alignment only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully correct</td>
<td>19.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Correct end points</td>
<td>1.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
</tr>
<tr>
<td>One correct end point</td>
<td>50.8</td>
<td>77.9</td>
<td>72.3</td>
<td>80.9</td>
</tr>
<tr>
<td>Wrong but near</td>
<td>6.6</td>
<td>18.3</td>
<td>16.2</td>
<td>16.1</td>
</tr>
<tr>
<td>Wrong</td>
<td>1.0</td>
<td>1.7</td>
<td>1.5</td>
<td>2.4</td>
</tr>
<tr>
<td><strong>Most favorable alignment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully correct</td>
<td>51.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Correct end points</td>
<td>11.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
</tr>
<tr>
<td>One correct end point</td>
<td>14.5</td>
<td>85.6</td>
<td>84.6</td>
<td>80.9</td>
</tr>
<tr>
<td>Wrong but near</td>
<td>0.9</td>
<td>11.3</td>
<td>4.7</td>
<td>16.1</td>
</tr>
<tr>
<td>Wrong</td>
<td>1.0</td>
<td>1.1</td>
<td>0.8</td>
<td>2.4</td>
</tr>
</tbody>
</table>
LASER Features

- ... maps seeds using standard aligner (e.g. BWA)
- ... has high sensitivity
- ... aligns regular and breakpoint-covering reads
- ... performs outperforms other read mappers for indels $> 20$bp
- ... produces BAM files

Integration into existing pipelines
Use BAM file for downstream analyses; e.g. GATK genotyping
MATE-CLEVER =
CLEVER + LASER + (trio-aware) genotyping

Reference
Genome of the Netherlands (GoNL)

Overview

- Sequencing the (whole) genomes of 250 Dutch families:
  - 231 trios: mother, father, child
  - 8 quartets: mother, father, dizygotic twins
  - 11 quartets: mother, father, monozygotic twins
- → 769 individuals in total

Take family structure into account while genotyping
MATE-CLEVER Workflow

**Input:** BWA alignments

1. Run CLEVER on each individual
2. Pool predictions, extract regions around indels
3. Realign these regions using LASER
4. Pool possible SNP and indel positions
5. Mark indels predicted by CLEVER and LASER
6. Recalibrate scores of LASER alignments
7. Perform genotyping based on recalibrated LASER alignments
## Genotyping

### Given
The number $n_C$ of **contradicting** reads and the number $n_S$ of **supporting** reads

### Sought
The **genotype distribution**, i.e. the probabilities of the deletion being “absent”, “heterozygous”, or “homozygous”

Let $\mathcal{B}_{n,p}$ be a binomial distribution

- $p_{\text{absent}} \propto \mathcal{B}_{n,p}(n_S)$, $p = p_{FP}$
- $p_{\text{heterozygous}} \propto \mathcal{B}_{n,p}(n_S)$, $p = 0.5p_{FP} + 0.5(1 - p_{FN})$
- $p_{\text{homozygous}} \propto \mathcal{B}_{n,p}(n_S)$, $p = 1 - p_{FN}$
Bayesian Genotyping

**One individual**
\[ p_g \propto p_g^{\text{prior}} \cdot p_g^{\text{ins}} \cdot p_g^{\text{split}} \quad \text{for} \quad g \in \{\text{absent, heterozygous, homozygous}\} \]

**Trio (mother, father, child)**
\[ p_{fgh} = p_{fgh}^{\text{Mendel}} \cdot p_f^{\text{mother}} \cdot p_g^{\text{father}} \cdot p_h^{\text{child}} \]

**Exception**
De novo mutations, see next slide
Special case: *de novo* deletions

Sometimes *new mutations* arise in the child (i.e. mutations not present in mother or father)

**Conditions**

- $p_{\text{child \ absent}} < T$
- $1 - p_{\text{mother \ absent}} < T$
- $1 - p_{\text{father \ absent}} < T$

for a user-set threshold $T$.

*De novo* mutations and the mechanisms causing them are most interesting biologically.
## Performance on Venter Trio (30x): 10–49 bp

<table>
<thead>
<tr>
<th></th>
<th>Overall Recall</th>
<th>Family Precision</th>
<th>Individual Precision</th>
<th>Genotype Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length range 10–29</strong> (25,678 true deletions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATE-CLEVER</td>
<td>22.2</td>
<td>91.0</td>
<td>90.3</td>
<td>83.4</td>
</tr>
<tr>
<td>GATK</td>
<td>69.9</td>
<td>89.8</td>
<td><strong>90.6</strong></td>
<td><strong>88.3</strong></td>
</tr>
<tr>
<td>PINDEL</td>
<td><strong>82.7</strong></td>
<td>90.4</td>
<td>86.8</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Length range 30–49</strong> (3,170 true deletions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATE-CLEVER</td>
<td><strong>69.4</strong></td>
<td>86.1</td>
<td>87.7</td>
<td>84.6</td>
</tr>
<tr>
<td>GATK</td>
<td>23.2</td>
<td><strong>91.3</strong></td>
<td><strong>90.7</strong></td>
<td><strong>88.5</strong></td>
</tr>
<tr>
<td>PINDEL</td>
<td>58.8</td>
<td>73.8</td>
<td>77.7</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Performance on Venter Trio (30x): 50–249 bp

<table>
<thead>
<tr>
<th></th>
<th>Overall Recall</th>
<th>Family Precision</th>
<th>Individual Precision</th>
<th>Genotype Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length range 50–99</strong> (1,854 true deletions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATE-CLEVER</td>
<td><strong>55.5</strong></td>
<td><strong>74.5</strong></td>
<td><strong>79.5</strong></td>
<td>76.9</td>
</tr>
<tr>
<td>GATK</td>
<td>0.6</td>
<td>73.3</td>
<td>77.8</td>
<td><strong>77.8</strong></td>
</tr>
<tr>
<td>PINDEL</td>
<td>37.2</td>
<td>55.9</td>
<td>69.0</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Length range 100–249</strong> (1,137 true deletions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATE-CLEVER</td>
<td><strong>42.4</strong></td>
<td><strong>63.3</strong></td>
<td><strong>72.6</strong></td>
<td><strong>70.2</strong></td>
</tr>
<tr>
<td>GATK</td>
<td>0.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PINDEL</td>
<td>20.1</td>
<td>58.7</td>
<td>70.7</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Structural Variation Discovery and Genotyping
Example: 7466bp deletion
Acknowledgments

- **Alexander Schönhuth**, CWI
- Markus Bauer, formerly Illumina, now Boehringer
- Stefan Canzar, formerly CWI, now Johns Hopkins
- Ivan Costa, formerly Recife, Brazil, now RWTH Aachen
- Iman Hajirasouliha, Brown University
- Gunnar Klau, CWI
- Alexander Schliep, Rutgers
- Ole Schulz-Trieglaff, Illumina

Software

Availability: http://clever-sv.googlecode.com
GNU Public License