RGASP workshop

Christian Iseli
November 2009

LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Analysis pipeline

- Filter input
- Classify pairs
- Spliced alignment
- Group alignments
- Generate output

LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Filter input

- Low quality nucleotides -> N
- Trim Ns on both ends
- Limit multiplicative effect
- Detect and eliminate repeats

Classify pairs (1)

- Use fetchGWI
- Find unique exact matches
- Verify paired reads distance and orientation

LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Classify pairs (2)

	Pairs 2 matched	Pairs 1 matched	Pairs 0 matched	Pairs non-canonical	Singlet Mapped	Singlet No match	Total
GMr1	7.4 M	19.0 M	58.3 M	1.7 M	0.06 M	0.2 M	86.6 M
GMr2	5.9 M	12.2 M	28.7 M	1.7 M	0.03 M	0.3 M	48.8 M
Kr1	5.8 M	18.5 M	45.5 M	1.4 M	0.04 M	0.1 M	71.4 M

LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Classify pairs (3)

- Use megablast on unmapped pairs
- Verify distance and orientation

	GMr1	GMr2	Kr1
Pairs 0 matched	58250606	28657924	45522710
megablast	12412160	5801018	11162500

LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Spliced alignment (1)

- Use paired information to restrict search area
- Use SIBsim4

Spliced alignment (2)

```
>chr|NC_000001|NC_000001.10 Chromosome 1; LEN=160430423
>tag|4_8_1524_762.1|- clone 4_8_1524_762; LEN=76

160295414-160295442  (1-29)   100% <- (GT/AG) 24
160302238-160302284  (30-76)   100%

        0     .    :    .    :    .    :    .    :    .    :
160295414 ACCAGGGGACAGGTTTTTTTTCCTCAGACCTT...CACCAGAAATATCCC
          |||||||||||||||||||||||||||||<<<...<<<||||||||||||
        1 ACCAGGGGACAGGTTTTTTTTCCTCAGAC         CAGAAATATCCC

       50     .    :    .    :    .    :    .
160302250 AAACGCGCACAGTCTGGTCCAGGCTGGCTGATACT
          |||||||||||||||||||||||||||||||||||
       42 AAACGCGCACAGTCTGGTCCAGGCTGGCTGATACT
```

Group alignments (1)

- Use 'tromer' to combine alignments
- Use paired reads information
- Predict transcripts

LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Group alignments (2)

```
>map|NC_000001_37|NC_000001.10|+ Chromosome 1 541739..567794
H0 [541739] 0,1
 =E0=> D0 1,7
D0 [541752] 1,1
 H0 1,7 =E0=>
 -I0-> A0 1,7
A0 [567506] 1,1
 D0 1,7 -I0->
 =E1=> T0 2,7
T0 [567794] 1,0
 A0 2,7 =E1=>

E0 541739..541752 0,1
 t:4_57_813_1318.2 6..19 (541739..541752) 85%
E1 567506..567794 0,2
 L:4_50_1286_288 1..253 (567542..567794) 100%
 t:4_57_813_1318.2 20..76 (567506..567561) 98%
I0 541753..567505 1
 t:4_57_813_1318.2 19..20 GT/AG 7
```

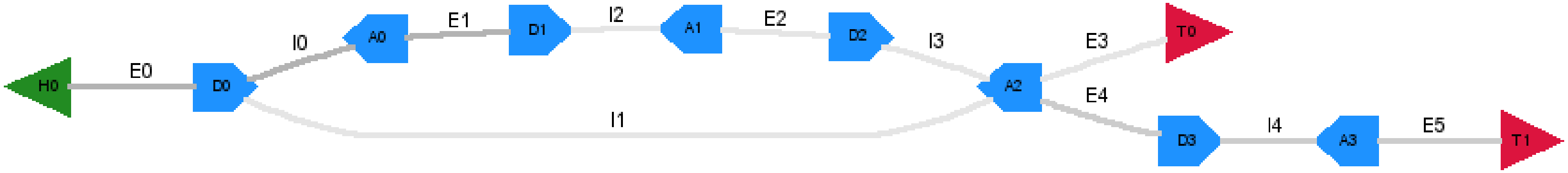
LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Group alignments (3)

	Reads	Pairs	Graphs
GM12878 r1	35,210,742	18,189,056	835,017
GM12878 r2	22,208,026	11,479,864	497,331
GM12878	57,418,768	29,501,799	895,252
K562 r1	32,250,482	16,661,568	660,902

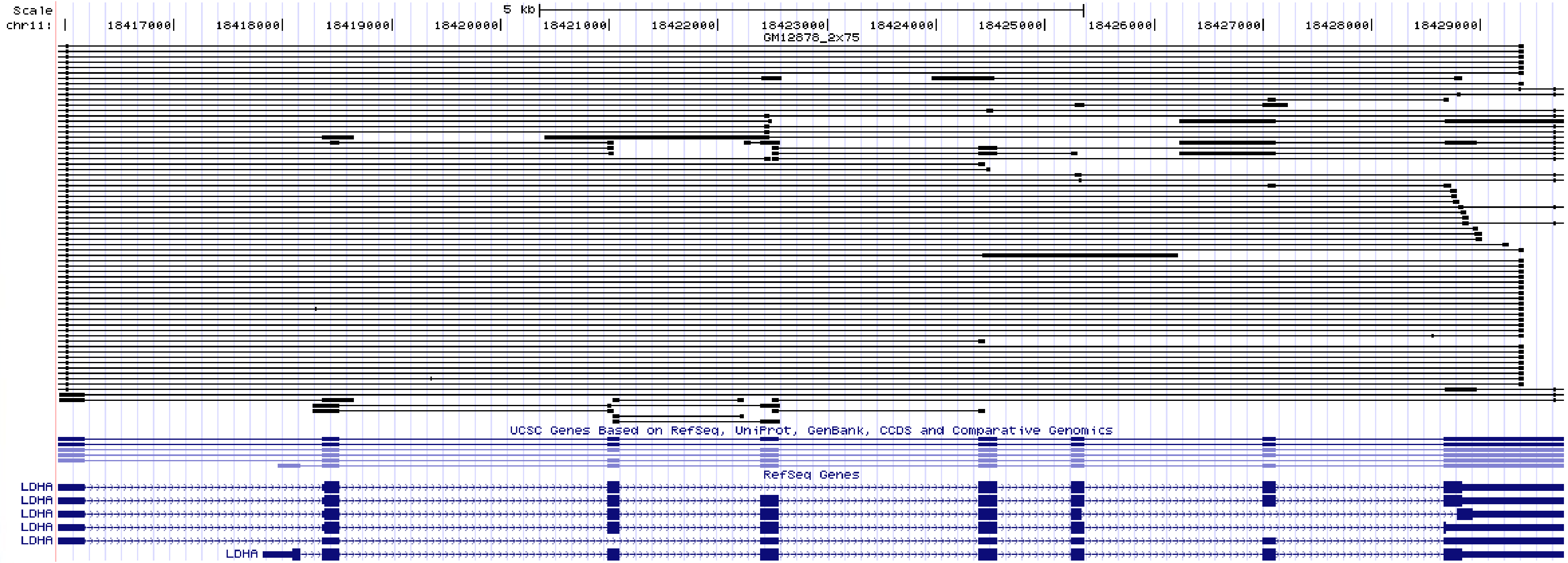
Generate output (1)



LUDWIG INSTITUTE FOR CANCER RESEARCH

SIB

Generate output (2)



LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB

Acknowledgments

Ludwig Institute

- *Armand Valsesia*
- *Brian Stevenson*

SIB

- *Vital-IT*
- *Ioannis Xenarios*
- *Heinz Stockinger*

EPFL Life Sciences

- *Jacques Rougemont*
- *Philipp Bucher*
- *Gregory Lefebvre*
- *Giovanna Ambrosini*
- *Morteza Ahmadi*
- *Dimitrieva Slavica*

LUDWIG
INSTITUTE
FOR
CANCER
RESEARCH

SIB