ADTEx User Guide

Kaushalya C. Amarasinghe

kaushalyac@tgmail.com

24 September 2013

1. Introduction

ADTEx is a copy number variation (CNV) detection and zygosity detection tool in cancer exome sequencing data. It uses normalized depth of coverage (DOC) ratios among tumour and its matched normal sample to predict CNV. Also utilizes B allele frequencies of tumour sample at heterozygous loci in control sample to predict zygosity states of CNV regions.

2. Requirements

- R2.15 – The main analysis of CNV using Hidden Markov Model (HMM) and discrete wavelet transformation (DWT) are done using R statistical language.
  http://www.r---project.org/
- "wmtsa" R package
  http://cran.r-project.org/web/packages/wmtsa/index.html
- "DNAcopy" R package
  http://www.bioconductor.org/packages/release/bioc/html/DNAcopy.html
- Python 2.7+ - Easy to use tool is developed mainly using python.
  http://www.python.org/
- BEDTools
  http://code.google.com/p/bedtools/

3. Input file types

1. Tumour/matched normal data

- BAM format of the sequence alignemnts
  or
- Tab delimited file with coverage information
  These can be created using BEDTools command:

  coverageBed –abam <BAM input> -b <targets in bed format> -d > <output.file>

  Variations of input files for the above command can be found in BEDTool's documentation. However, the option (-d) should be used.

The output of this command should look like the following without headers:

<chr>	<start>	<end>	<position>	<# of reads>
1	6709	7001	1	100
1	6709	7001	2	105
1	6709	7001	3	110
1	6709	7001	4	115
1	6709	7001	5	105

2. Target definition files
   This will have the targets of the exome capture in BED format (http://genome.ucsc.edu/FAQ/FAQformat)
   All chromosome definitions should be in the same format as the BAM or DOC file (i.e. '1' or 'chr1' for chromosome 1).

3. B allele frequencies file
   This is a tab delimited file with following essential headers.

- chrom – chromosome name (same format as in BED or BAM file)
- SNP_loc – location of the SNP
- control_BAF – B allele frequency (BAF) at each SNP in control sample
- tumor_BAF – B allele frequency (BAF) at each SNP in tumor sample
- control_doc – Total read count at each SNP in control sample
- tumor_doc – Total read count at each SNP in tumour sample

4. Output format

1. Tab delimited 'cnv.result' file
   The result will have copy number prediction for each exon. This will have other information such as chromosome, exon start, exon end, DOC ratio after normalization and segment mean in tab delimited format.

2. Plots of CNV results
   If plotting option is specified in the command, this will plot DOC ratios for each chromosome separately as '.png' file. The segments will be coloured based on the predicted copy number.

3. Tab delimited 'zygosity.res' file
   This has zygosity state of each heterozygous SNP locus.

5. Options available in ADTEx

--normal / -n <input file>	Input file for matched normal [required]
--tumor / -t <input file>	Input file for tumor [required]
--bed / -b <input file>	Target definitions [required]
--out / -o <output folder>	Output folder path [required] (Programme will create a folder with this name, so it must not be a name of a existing folder)
--DOC	Specify whether tumor and matched normal input give coverage information [required only when input in DOC format]
--ploidy	Common ploidy of the tumor [optional]. Default: 2
--estimatePloidy	This will estimate base ploidy when --ploidy option is not provided. However, --baf must be provided to execute this command [optional]. Default : False
--plot / -p	If specified, the output plots will be generated. [optional]
--minReadDepth	Exons having avg. reads <minReadDepth in the control sample will not be analysed [optional]. Default: 10
--baf	If specified BAFs will be used to predict zygosity states and correction of copy number predictions [option al]. The B allele frequency file should be specified with this option.

6. Running ADTEx

1. With BAM inputs

`python ADTEx.py --normal` *normal_sample.BAM* `--tumor` *tumor_sample.BAM* `--bed` *target.bed* `--out` *output_folder*

2. With coverage files

`python ADTEx.py --normal` *normal_sample* `--tumor` *tumor_sample* `--bed` *target.bed* `--out` *output_folder* `--DOC`