

**From:** Wen Yao <ywhzau@gmail.com>  
**Time:** 2014/4/14 9:55  
**To:** 'Radesh Nattamai Malli'  
**Subject:** Re: Re: Re: Re: Re: Intansv - Reference paper  
**Attach:** Supplementary.pdf

Dear Radesh,

I'm really sorry for the late response.

The paper for intansv is not yet published because I was busy with another project and I still want to add some new features to intansv.

There are default parameters for all the functions in the intansv package. You can see the help manual for each function in R using ?+(function name) (for example ?readDelly).

There are two default parameters for readCnvator: (regSizeLowerCutoff=100, regSizeUpperCutoff=1000000).

3 default parameters for readDelly: (regSizeLowerCutoff=100, regSizeUpperCutoff=1000000, readsSupport=3)

3 default parameters for readPindel: (regSizeLowerCutoff=100, regSizeUpperCutoff=1000000, readsSupport=3)

3 default parameters for readSvseq: (regSizeLowerCutoff=100, regSizeUpperCutoff=1000000, readsSupport=3)

4 default parameters for readBreakDancer: (regSizeLowerCutoff=100, regSizeUpperCutoff=1000000, readsSupport=3, scoreThreshold=0.5)

As you can see, the result of CNVator were filtered by SV (structural variation) size only. the result of Pindel, Svseq2, Delly

There is one additional step:

The output of programs like BreakDancer or Pindel usually contain overlapped predictions for the same SV. These overlapped

The default parameters for methodsMerge are: overLapPerDel = 0.8, overLapPerDup = 0.8, overLapPerInv = 0.8, numMethod

The detail of methodsMerge:

Predictions of different programs were first clustered together if they have reciprocal coordinate overlap of more than 80%

I also sent you an attachment (with two figures and legends) for the explanation of processes for the filtering step and the

Handle redundancy by methodsMerge:

If redundant SVs by different methods have reciprocal coordinate overlap of more than 80% (default value), they were merged

For "handle frequencies of the sv calls", I am not quite sure what do you mean.

I hope this is helpful to you. And if you have further questions, please don't hesitate to contact me.

Best regards,

Wen

**From:** Radesh Nattamai Malli [<mailto:>]  
**Time:** 2014/4/11 12:39  
**To:** Wen Yao  
**Subject:** Fwd: Re: Re: Re: Re: Re: Intansv - Reference paper

Dear Mr. Yao,

I hope you had a chance to read my below email. I kindly request your input and suggestions regarding the below email. Once again I thank you for your time. Looking forward to hearing back from you.

Regards,  
Radesh

----- Forwarded message -----

**From:** Radesh Nattamai Malli <>  
**Date:** Wed, Apr 9, 2014 at 3:17 AM  
**Subject:** Re: Re: Re: Re: Re: Re: Intansv - Reference paper  
**To:** Wen Yao <[ywhzau@gmail.com](mailto:ywhzau@gmail.com)>

Dear Mr. Yao,

Thank you for the update for the new version of intansv, I have updated it to the current version and am using it. The update is working well especially in methods merge where i do not have to give the names of the tools which am not using.

For cnvator the command has the default values.

```
C12_cnvator <- readCnvator(dataDir="/home/rnattamaimallip/work/C12_cnvator", regSizeLowerCutoff=100, regSizeUpperCutoff=1000000)
```

-----  
Please let me know what are the default values used when executing readDelly, readBreakDancer and methodsMerge.

```
C12_delly <- readDelly("/home/rnattamaimallip/work/C12_delly")
```

```
C12_breakdancer <- readBreakDancer("/home/rnattamaimallip/work/C12_bd_copy/combined.sv")
```

```
sv_all_methods <- methodsMerge (breakdancer = C12_breakdancer, cnvator = C12_cnvator, delly = C12_delly)
```

Kindly explain how does the methods merge work in detail, like what are the default values ? How does it handle frequencies of the sv calls ? How does the merging work? How does it handle redundancy ?

I would appreciate it very much if you could explain the workings. Have you published the paper? Can you send me the journal name and link. Looking forward to hearing back from you.

Thank you and have a great day

Radesh

On Mar 1, 2014 8:56 AM, "Wen Yao" <[ywhzau@gmail.com](mailto:ywhzau@gmail.com)> wrote:

Dear Radesh,

I have released a new version of intansv which now support read in the results of two new tools: Lumpy and SoftSearch. Several functions had also been modified a little to make them more robust.

For the predictions of DELLY. I think you can modify the results of DELLY before give in to intansv since readDelly only cares about lines with "Deletions/Duplication/Inversion" in the output of DELLY. By only keep lines with "Deletions/Duplication/Inversion" of the output of DELLY and then give it to readDelly, will accelerate readDelly quite a lot.

Best regards,  
Wen

**From:** Radesh Nattamai Malli [mailto:]

**Time:** 2014/2/27 5:51

**To:** Wen Yao

**Subject:** Re: Re: Re: Re: Re: Intansv - Reference paper

Dear Mr. Yao,

I once thank you and really appreciate your input and emails. As you probably guessed I am working with NGS data for around 100 individuals.

In regards to the breakdancer, I did combine all the results to a single file and was able to execute the command. I also understand the time constraint in processing a large amount of data in regards to Delly. It is fine as long as the results are coming.

I checked your documentation for annotation and geneannotation, and the examples require further information, I am trying to find out about generating a .gff file or a genomeannotation files and sv annotation file for the results I get. if there is sequence of steps for the creation of each of the file, please let me know.

Another question,

How long are the results stored in memory after a command has been executed. I found that once a command has been executed and read, it does not stay in memory and i have to run the command again to read the data. That leads to question of how can I export data as .txt file from R. I did try sink, cat and filConn to export the data and I was not able to get all the 8 variables in a txt file to analyze it later. Similarly I would also want to import the .txt file as input to further annotate or process it. Would it be possible to do that. I would really appreciate your input and suggestions. As there is no documentation for this I am requesting your input.

once again Thank you  
Regards  
Radesh

On Mon, Feb 24, 2014 at 8:34 PM, Wen Yao <[ywhzau@gmail.com](mailto:ywhzau@gmail.com)> wrote:

Dear Radesh,

Currently, readBreakDancer doesn't support read in a directory. With a directory of 2GB size, readDelly may take too long time.

I am wondering about your experiment. Did you perform BreakDancer and DELLY with the NGS data of a population of tens or hundreds of individuals? Did you want to read in them all at once?

If readBreakDancer does need to read in a directory, I will make it to do that. I will also try to make readDelly more fast.

If you have any suggestions, please contact me.

Best regards,  
Wen Yao

**From:** Radesh Nattamai Malli [mailto:]  
**Time:** 2014/2/24 15:30  
**To:** Wen Yao  
**Subject:** Re: Re: Re: Re: Intansv - Reference paper

Dear Mr. Yao,

In regards to your previous email, I wish to request your input for executing breakdancer.

Below is the logs of the commands for executing breakdancer, I am able to execute individual files but when i input the directory which has the files i get an error message which is listed below. any input or suggestions to work around this issue. I would much appreciate your comments.

```
breakdancer <- readBreakDancer("/home/rnattaimallip/work/readBreakDancer", scoreCutoff=60,
readsSupport=3,regSizeLowerCutoff=100, regSizeUpperCutoff=1000000)
Error in read.table(breakDancer, colClasses = bdColClass) :
  no lines available in input
> breakdancer <- readBreakDancer("/home/rnattaimallip/work/readBreakDancer/*.rv.sv", scoreCutoff=60,
readsSupport=3,regSizeLowerCutoff=100, regSizeUpperCutoff=1000000)
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file '/home/rnattaimallip/work/readBreakDancer/*.rv.sv': No such file or directory
> breakdancer <- readBreakDancer("/home/rnattaimallip/work/readBreakDancer")
Error in read.table(breakDancer, colClasses = bdColClass) :
  no lines available in input
> breakdancer <- readBreakDancer("/home/rnattaimallip/work/readBreakDancer/")
Error in read.table(breakDancer, colClasses = bdColClass) :
  no lines available in input
> breakdancer <- readBreakDancer("/home/rnattaimallip/work/readBreakDancer/01C08034.rv.sv") ----- able to
open individual dataset- but not the directory - all the files are checked for spaces and revised.
> str(breakdancer)
List of 2
$ del:'data.frame': 3769 obs. of 4 variables:
..$ chromosome: chr [1:3769] "chr1" "chr1" "chr1" "chr1" ...
..$ pos1 : num [1:3769] 869448 1120144 1866382 2038195 2390932 ...
..$ pos2 : num [1:3769] 870298 1120265 1867008 2038370 2391047 ...
..$ size : num [1:3769] 851 121 626 175 115 172 163 127 246 109 ...
$ inv:'data.frame': 93 obs. of 4 variables:
..$ chromosome: chr [1:93] "chr1" "chr1" "chr1" "chr1" ...
..$ pos1 : num [1:93] 16949717 39962026 44059827 80795217 81660526 ...
..$ pos2 : num [1:93] 17279464 40249696 44060038 80795395 81661208 ...
```

..\$ size : num [1:93] 329747 287670 211 178 682 ...

The dataset for delly is 2.0 GB and it has 123 files in the directory, it took close to 2 hrs for the program to run, is that normal running time, have you had a chance to check this timing issue.

I am finding this tool really useful and am testing it a lot. Looking forward to hearing your comments and suggestions.

Warm regards,  
Radesh

On Thu, Feb 20, 2014 at 2:24 AM, Wen Yao <[ywhzau@gmail.com](mailto:ywhzau@gmail.com)> wrote:

Dear Radesh,

You are welcome to give any suggestions and feedbacks.

Currently intansv only deal with deletions, inversions and duplication since only few programs could predict translocations and translocations are more complex to predict. Besides, the output of translocations are different from each program. BreakDancer doesn't report the direction of translocation while delly does. And I am considering to support more SV prediction programs (like lumpy and softsearch). Translocations might be supported in the future release of intansv.

Best regards,

Wen Yao

**From:** Radesh Nattamai Malli [mailto:]  
**Time:** 2014/2/20 14:05  
**To:** Wen Yao  
**Subject:** Re: Re: Re: Intansv - Reference paper

Dear Mr. Yao,

Thank you for your kind response and clarifications. I was able to follow and execute the commands as suggested.

I hope you don't mind if I can ask you about translocations(ctx, itx, jmp files in delly etc.) not being read by the package even if the input file has them. I shall use the package and get back to you if I have further clarifications.

Thank you once again.  
Radesh

On Feb 19, 2014 2:25 AM, "Wen Yao" <[ywhzau@gmail.com](mailto:ywhzau@gmail.com)> wrote:

Dear Radesh,

Thanks for your interest in intansv.

The example data for the intansv package were packed with it. You can find the example data where intansv is installed in your system. First, find the path where intansv is installed:

```
> find.package("intansv")
```

```
[1] "C:/Program Files/R/R-3.0.1/library/intansv"
```

So, the example data is here:

```
C:\Program Files\R\R-3.0.1\library\intansv\extdata
```

In fact, the command "system.file" is an R function. You can find the help document for it by typing "?system.file" in R. The example I provide:

```
breakdancer <- readBreakDancer(system.file("extdata/ZS97.breakdancer.sv",  
                                           package="intansv"))
```

The file [ZS97.breakdancer.sv](#) is packed with the intansv package. In my system, it is in "C:\Program Files\R\R-3.0.1\library\intansv\extdata".

But in your system, it may not be there. The "system.file" is to find the path of this file.

For your data, you should use:

```
breakdancer <- readBreakDancer("/home/rnattamaimallip/work/readBreakDancer/01C06309.ctx.sv")
```

```
delly <- readDelly("/home/rnattamaimallip/work/readDelly/")
```

Since you have the full path of these files.

For the 'methodsMerge' command, you may use this:

```
methodsMerge(breakdancer=breakdancer, pindel=pindel, cnvnator=cnvnator, delly=delly, svseq=svseq)
```

if you have the output of all the 5 programs.

Or you may use this:

```
methodsMerge(breakdancer=breakdancer, pindel=NULL, cnvnator=cnvnator, delly=delly, svseq=svseq)
```

if you have output of 4 programs (no pindel).

Or you may use this:

```
methodsMerge(breakdancer=breakdancer, pindel=NULL, cnvnator=cnvnator, delly=delly, svseq=NULL)
```

if you have output of 3 programs (no Pindel and svseq).

.....

I will improve the manual for each of the function and all the documentations as soon as possible.

I will also improve these functions to make them more friendly.

If you encounter any problems using intansv in future, please contact me.

Best regards,

Wen Yao

**From:** Radesh Nattamai Malli [mailto:]

**Time:** 2014/2/19 13:36

**To:** Wen Yao

**Subject:** Re: Re: Intansv - Reference paper

Dear Mr. Yao,

Thank you for response. I would appreciate it very much if you could let me know how to view the example data for the intansv package.

I am having difficulties in loading breakdancer, below is the log for your reference,

```
> breakdancer <- readBreakDancer(system.file  
("/home/rnattamaimallip/work/readBreakDancer/01C06309.ctx.sv", package="intansv"))  
Error in read.table(breakDancer, colClasses = bdColClass) :  
  no lines available in input  
In addition: Warning message:  
In file(file, "rt") :  
  file("") only supports open = "w+" and open = "w+b": using the former  
  
> delly <- readDelly(system.file("/home/rnattamaimallip/work/readDelly/", package="intansv"))  
> str(delly)  
List of 3  
 $ del: NULL  
 $ dup: NULL  
 $ inv: NULL
```

I am able to read the delly files only by using the below command, similarly for CNVnator files.

```
readDelly(dataDir="/home/rnattamaimallip/work/readDelly",  
regSizeLowerCutoff=100, regSizeUpperCutoff=1000000, readsSupport=3)
```

```
$del - 8185 sv determined/predicted
$dup - 621 sv determined/predicted
$inv - 198 sv determined/predicted
```

is there any particular reason why using system.file gives null results.

when I try to merge the files below is the log

```
methodsMerge( cnvnator, delly, others=NULL,
+ overLapPerDel = 0.8, overLapPerDup = 0.8, overLapPerInv = 0.8,
+ numMethodsSupDel = 2, numMethodsSupDup = 2, numMethodsSupInv = 2)
Error in methodsMerge(cnvator, delly, others = NULL, overLapPerDel = 0.8, :
  object 'cnvnator' not found
```

```
sv_all_methods <- methodsMerge(cnvator,delly)
Error in methodsMerge(cnvator, delly) : object 'cnvnator' not found
> sv_all_methods <- methodsMerge(breakdancer,delly)
Error in methodsMerge(breakdancer, delly) :
  argument "cnvnator" is missing, with no default
> sv_all_methods <- methodsMerge(breakdancer, cnvnator, delly)
Error in methodsMerge(breakdancer, cnvnator, delly) :
  object 'cnvnator' not found
```

I would really appreciate your input and suggestions. If you could provide more reference/documentation to your package, it would be of great help. Looking forward to hearing back from you at the earliest,  
Thank you in advance.

Radesh

On Fri, Feb 7, 2014 at 2:42 AM, Wen Yao <[ywhzau@gmail.com](mailto:ywhzau@gmail.com)> wrote:

Dear Radesh,

I am really sorry for the late response. Thanks for your interest in intansv.  
I came back home for the Chinese New Year and didn't have access to the internet.  
Currently the paper of intansv has not yet been published. I want to add some other features to intansv.  
I will try to deposit the manuscript to [arxiv.org](http://arxiv.org) before sent for publication. Then you can download it.

If you have any suggestions, please contact me.

Best regards,  
Wen Yao

**From:** Radesh Nattamai Malli [mailto:]  
**Time:** 2014/1/30 7:56  
**To:** [ywhzau@gmail.com](mailto:ywhzau@gmail.com)  
**Subject:** Intansv - Reference paper

Dear Mr. Yao,

I came across your intansv package in Bioconductor site. It is a really interesting and quite useful tool. I can't seem to find the journal reference paper. I would appreciate it very much if you could send me the link for downloading the paper.

Thanks in advance

Sent from Samsung Mobile  
Radesh NMP

--

Radesh P Nattamai Malli

--

Radesh P Nattamai Malli

--

Radesh P Nattamai Malli

--

Radesh P Nattamai Malli