The Author(s) *BMC Genomics* 2016, **17**(Suppl 5):496
DOI 10.1186/s12864-016-2824-x

BMC Genomics

RESEARCH Open Access

SPAI: an interactive platform for indel analysis

Mohammad Shabbir Hasan and Liqing Zhang*

From 11th International Symposium on Bioinformatics Research and Applications (ISBRA '15)
Norfolk, VA, USA. 7–10 June 2015

Abstract

Background: Insertions and Deletions (Indels) are the most common form of structural variation in human genome. Indels not only contribute to genetic diversity but also cause diseases. Therefore assessing indels in human genome has become an interesting topic to the research community. This increasing interest on indel calling research has resulted into the development of a good number of indel calling tools. However, all of these tools are command line based and require expertise from Computer Science (CS) to execute them which makes it challenging for researchers from non-CS background.

Methods: In this paper, we describe an interactive platform named SPAI which stands for Single Platform for Analyzing Indels.

Results: Being a Graphical User Interface (GUI) tool, SPAI facilitates users to run several popular indel calling tools and perform several analyses on the indel calling results without knowing any command line programming.

Conclusions: SPAI is written in Java and tested in Linux operating system.

Background

Single Nucleotide Polymorphisms (SNPs) constitute the major portion of genetic variations that happen in human genome. However, recent studies show that insertion and deletions which are collectively known as indels also contribute to genetic diversity dramatically [1, 2]. Being a structural variant, indels can alter human phenotype and therefore can cause several kinds of diseases [3, 4]. For example, Cystic Fibrosis, one of the most common genetic diseases in humans, is caused due to the deletion of 3 base pairs (bps) which leads to the elimination of a single amino acid from the encoded protein [5]. Similarly insertion of base pairs in the DNA sequence results change in gene function that cause diseases like Fragile X Syndrome [6], Mendelian disorders [7], Haemophilia [8], Neurofibromatosis [9], Muscular Dystrophy [10], and Cancer [11, 12]. In addition to causing diseases, indels within the promoter region influence gene expression and can be used to explain the difference in gene expression observed in different human [13] which apparently brings the use of indels as genetic markers in natural population [14].

With the introduction of Next Generation Sequencing (NGS) technology, now it is possible to sequence human genome at an unprecedented rate [1] and whole genome sequencing (WGS) is now possible at an individual level [15–19]. Whole genome sequencing has revealed numerous genetic variations that were not previously reported [20] and these variation profiles can also be used to predict ancestor's traits such as height, weight, appearance, and intelligence [1]. Therefore the idea of predicting the future health of individuals to design personalized medicine is rapidly approaching. However, accurate detection of genetic variation at an individual level is a key challenge in evolutionary genomic research. To accept this challenge, fortunately, a good number of indel calling tools have been developed so far [21, 22].

In recent time, many indel calling tools have been developed which are publicly available as well as popular among the researchers. Some of these tools include

* Correspondence: lqzhang@cs.vt.edu
Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA

Genome Analysis Tool Kit (GATK) [23, 24], VarScan [25, 26], Pindel [27], SAMtools [28], Dindel [29], Platypus [30], P-Dindel [21], SV-M [31], Stampy [32], PEMer [33], Hydra [34], BreakDancer [35], FreeBayes [36], and indelMINER [37]. A close look at these tools reveals that all of them are command line based [38]. Command line tools are very much useful for batch processing and they require a certain level of expertise from Computer Science (CS). Therefore, researchers from non-CS background such as Biology, Chemistry, and Microbiology, doing research on evolutionary genomics, often find it difficult to execute and explore different features of the tools by changing different parameters through command line.

Research by several usability labs reveal that the usability of a product can be significantly improved through the use of Graphical User Interface (GUI) [39] and providing the user with GUI for command line based bioinformatics tools became a success previously [40].

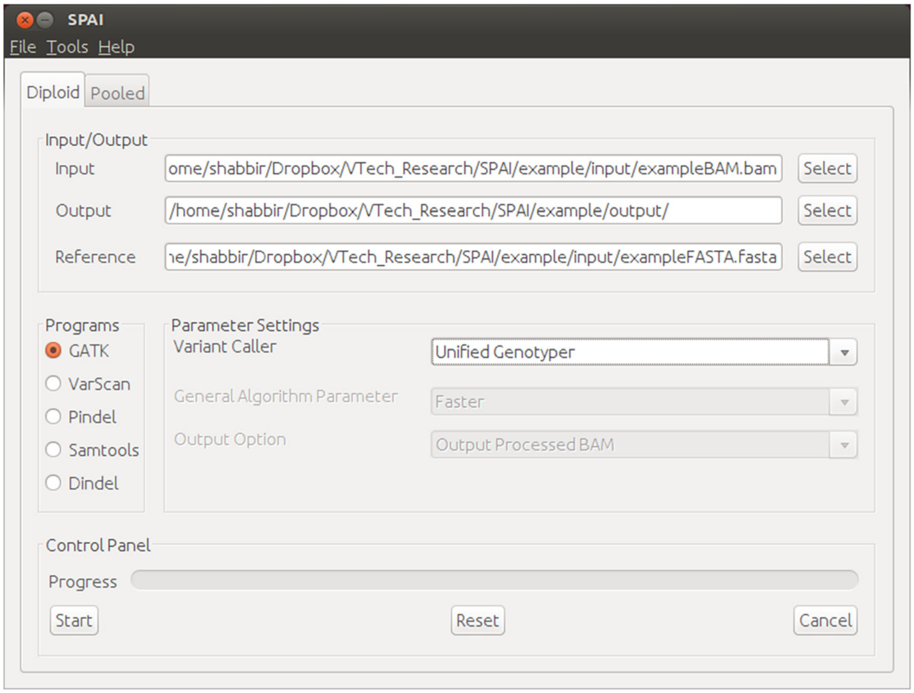
To come up with a solution to the usability problem of the indel calling tools, here we describe SPAI (Single Platform for Analyzing Indels). SPAI provides the user with a complete platform for indel research. In addition to running popular indel calling tools, through SPAI, user can also download alignment files (BAM files) form the 1000 Genome Project [41] to be used as input to these tools. Moreover, user can get coverage information of these alignment files and see the alignment files in a tabular format. In SPAI, user can also see the indel calling results in a tabular format to get a better insight of the called indels. For downstream analysis, SPAI lets the users to compare the results from various tools and visualize those comparisons using graphs and charts. Being an interactive tool, therefore, SPAI lets the user to perform necessary works of indel research without having any prior knowledge of command line programming.

Methods

SPAI which is written in Java comes with several features that are briefly described below.

Running different indel calling tools from GUI

Existing indel calling tools can be divided into four major categories: alignment based methods, split read methods, paired end read mapping methods, and haplotype based methods [22]. In the current version of SPAI, we include tools from two categories: alignment based methods and split read methods. From the alignment based methods category, SPAI includes GATK Unified Genotyper, VarScan, Dindel, and SAMtools. It also includes Pindel which belongs to the split read method category. When the user installs SPAI, these tools are also installed automatically. In the next release of SPAI, the following tools will be added: GATK Haplotype Caller, Platypus, FreeBayes, and indelMINER. Other tools from different categories will be added as the development of SPAI proceeds. Figure 1 shows the main GUI of SPAI. As we can see from this Figure, user just



Input ome/shabbir/Dropbox/VTech_Research/SPAI/example/input/exampleBAM.bam

Output /home/shabbir/Dropbox/VTech_Research/SPAI/example/output/

Reference ɿe/shabbir/Dropbox/VTech_Research/SPAI/example/input/exampleFASTA.fasta

● GATK
○ VarScan
○ Pindel
○ Samtools
○ Dindel

Variant Caller: Unified Genotyper

General Algorithm Parameter: Faster

Output Option: Output Processed BAM

Fig. 1 Main window of SPAI

needs to specify the inputs (alignment file and reference sequence file), output file location and which tool to run. Usually SPAI runs the selected indel calling tool using its default settings. However, SPAI allows advanced user to change different parameter of each of the tools and run those tools based on that settings.

Some of the programs require huge processing power to generate results in a reasonable time. Although the current version of SPAI is desktop based, we are now working on moving the processing step in a cloud based service so that enough processing power can be provided and in that case the computation time will no longer depend on the configuration of user's computer.

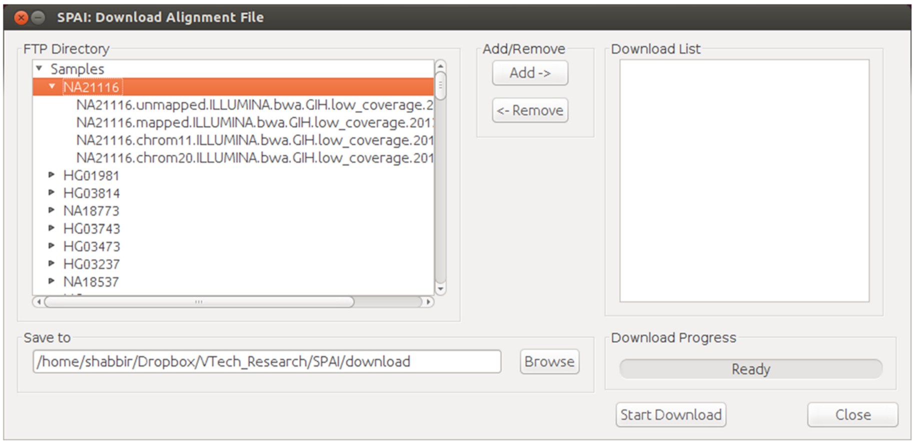
Download alignment files from the 1000 Genomes project

Input to the existing indel calling tools is the sequence alignment file which is usually available in BAM format. In most of the cases, the size of the BAM files is huge which can't be downloaded using the conventional downloader. To assist user in this case, an efficient downloader is integrated in SPAI which allows user to download single as well as multiple BAM files from the FTP server of the 1000 Genomes project. As shown in Fig. 2, the left panel of the downloader window shows the list of all human samples currently available in the 1000 Genomes project. From this list the user needs to select for which human and for which chromosome the alignment file is needed. After the selection is done, the file is added to the "Download List" as shown in the right panel of the downloader window. User can add multiple files to the downloader list or remove file from the list. After the selection is done, when the user hits the "Start Download" button, SPAI starts downloading the BAM file(s) and store it in the location specified by the user.

Sometimes it is inconvenient to store large BAM files in the local directory of the user's computer. Therefore, in the future release, SPAI will store the URLs of the alignment files in a text file in user's computer and will save the alignment file in a cloud based storage. Therefore while specifying the inputs, SPAI will allow user to put the URL of an alignment file instead of the physical location of that alignment file. It will also allow user to put external link to alignment files stored in different location other than the FTP server of the 1000 Genomes Project and in that case SPAI will fetch the alignment file and save it temporarily to the cloud based storage to be used during the execution time.

Comparing the results of different indel calling tools

User can compare the indel calling results produced by different tools which is a really useful feature for downstream analysis. In SPAI we provided a benchmark dataset [42] which contains 2 million small and large (length varies from 1 bp to 10,000 bps) indels found in the 24 chromosomes of 79 diverse humans. This dataset is considered to be the most reliable for indels in human genome and has been used as "gold standard" in other studies [22, 43, 44]. From the VCF (Variant Call Format) files that are generated by the tools as output, SPAI calculates the recall, precision, and F-measure of each tool after comparing their results with the benchmark dataset. User can see the comparisons as Graphs (shown in Fig. 3) which is really helpful to get an insight about the performance of the tools. Moreover, for the comparison purpose, SPAI allows user to supply results (in VCF format) from other tools that are not included in SPAI. This is really useful if the user want to assess the performance of a newly developed tool by comparing its result with existing tools as well as with "gold standard" indels.



Samples
NA21116
NA21116.unmapped.ILLUMINA.bwa.GIH.low_coverage.2
NA21116.mapped.ILLUMINA.bwa.GIH.low_coverage.201
NA21116.chrom11.ILLUMINA.bwa.GIH.low_coverage.201
NA21116.chrom20.ILLUMINA.bwa.GIH.low_coverage.201
HG01981
HG03814
NA18773
HG03743
HG03473
HG03237
NA18537

/home/shabbir/Dropbox/VTech_Research/SPAI/download

Fig. 2 Downloader window of SPAI

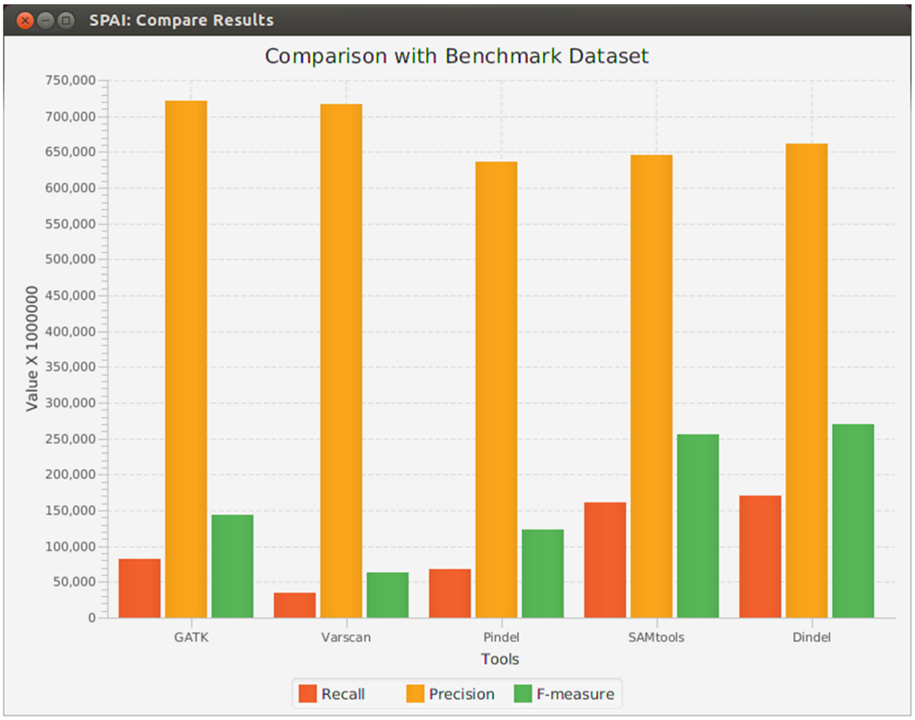


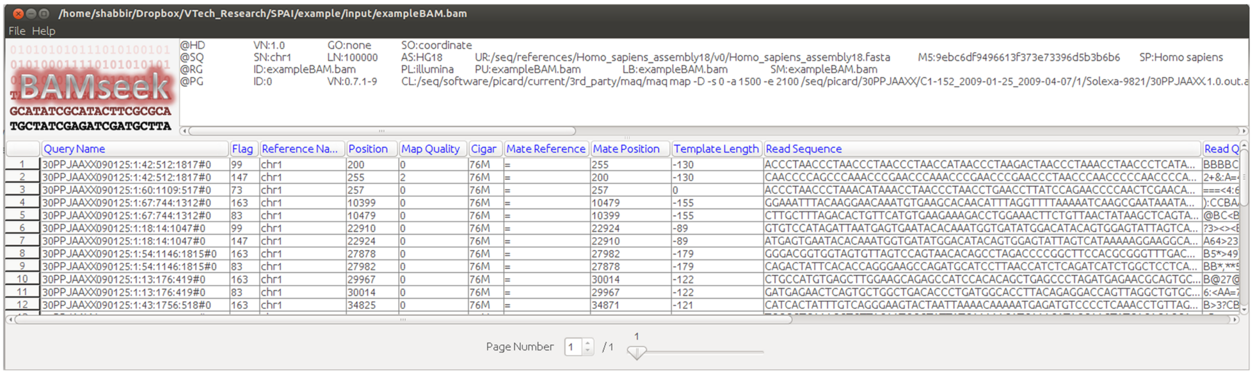
Fig. 3 Comparing the results of different tools with benchmark dataset

Displaying the alignment files and indel calling results in tabular format

Sometimes alignment files contain useful information about the reads (such as mapping quality, CIGAR string, type of the read etc.). However, the standard format of the alignment files is BAM which is a binary format and therefore, can't be opened using a text editor. Although SAMtools can convert BAM file to text format (SAM format), it is not convenient to open large SAM files using a text editor. To solve this problem, in SPAI, we include a third party tool called BAMSeek [45]. It can show large BAM files in a tabular format and user can get useful information about the alignment by hovering mouse to the corresponding column of the table. Similarly BAMSeek can also display large VCF files (output of the indel calling tools) in a tabular format from where user can get insight about the called indels. Figure 4 shows the tabular display of a BAM file.

Determining the depth of coverage

Depth of coverage is the average number of reads that represent a given nucleotide in the sequence. In most cases, high depth of coverage is desired for calling indels confidently [22]. SPAI allows users to determine whether an alignment file should be used as input for indel calling by calculating the coverage of that alignment file.


/home/shabbir/Dropbox/VTech_Research/SPAI/example/input/exampleBAM.bam
File Help
BAMseek
Page Number 1 /1
1

```
@HD    VN:1.0           GO:none        SO:coordinate
@SQ    SN:chr1          LN:100000      AS:HG18        UR:/seq/references/Homo_sapiens_assembly18/v0/Homo_sapiens_assembly18.fasta     M5:9ebc6df9496613f373e73396d5b3b6b6     SP:Homo sapiens
@RG    ID:exampleBAM.bam                PL:illumina    PU:exampleBAM.bam       LB:exampleBAM.bam        SM:exampleBAM.bam
@PG    ID:0             VN:0.7.1-9     CL:/seq/software/picard/current/3rd_party/maq/maq map -D -s 0 -a 1500 -e 2100 /seq/picard/30PPJAAXX/C1-152_2009-01-25_2009-04-07/1/Solexa-9821/30PPJAAXX.1.0.out.a
```

	Query Name	Flag	Reference Na...	Position	Map Quality	Cigar	Mate Reference	Mate Position	Template Length	Read Sequence	Read Q
1	30PPJAAXX090125:1:42:512:1817#0	99	chr1	200	0	76M	=	255	-130	ACCCTAACCCTAACCCTAACCCTAACCATAACCCTAAGACTAACCCTAAACCTAACCCTCATA...	BBBBC
2	30PPJAAXX090125:1:42:512:1817#0	147	chr1	255	2	76M	=	200	-130	CAACCCCAGCCCAAACCCGAACCCAAACCCGAACCCGAACCCTAACCCAACCCCCAACCCCA...	2+&:A=
3	30PPJAAXX090125:1:60:1109:517#0	73	chr1	257	0	76M	=	257	0	ACCCTAACCCTAAACATAAACCTAACCCTAACCTGAACCTTATCCAGAACCCCAACTCGAACA...	===<4:6
4	30PPJAAXX090125:1:67:744:1312#0	163	chr1	10399	0	76M	=	10479	-155	GGAAATTTACAAGGAACAAATGTGAAGCACAACATTTAGGTTTTAAAAATCAAGCGAATAAATA...	):CCBA
5	30PPJAAXX090125:1:67:744:1312#0	83	chr1	10479	0	76M	=	10399	-155	CTTGCTTTAGACACTGTTCATGTGAAGAAAGACCTGGAAACTTCTGTTAACTATAAGCTCAGTA...	@BC<B
6	30PPJAAXX090125:1:18:14:1047#0	99	chr1	22910	0	76M	=	22924	-89	GTGTCCATAGATTAATGAGTGAATACACAAATGGTGATATGGACATACAGTGGAGTATTAGTCA...	?3><><
7	30PPJAAXX090125:1:18:14:1047#0	147	chr1	22924	0	76M	=	22910	-89	ATGAGTGAATACACAAATGGTGATATGGACATACAGTGGAGTATTAGTCATAAAAAGGAAGGCA...	A64>23
8	30PPJAAXX090125:1:54:1146:1815#0	163	chr1	27878	0	76M	=	27982	-179	GGGACGGTGGTAGTGTTAGTCCAGTAACACAGCCTAGACCCCGGCTTCCACGCGGGTTTGAC...	B5*>49
9	30PPJAAXX090125:1:54:1146:1815#0	83	chr1	27982	0	76M	=	27878	-179	CAGACTATTCACACCAGGGAAGCCAGATGCATCCTTAACCATCTCAGATCATCTGGCTCCTCA...	BB*,**5
10	30PPJAAXX090125:1:13:176:419#0	163	chr1	29967	0	76M	=	30014	-122	CTGCCATGTGAGCTTGGAAGCAGAGCCATCCACACAGCTGAGCCCTAGATGAGAACGCAGTGC...	B@27@
11	30PPJAAXX090125:1:13:176:419#0	83	chr1	30014	0	76M	=	29967	-122	GATGAGAACTCAGTGCTGGCTGACACCCTGATGGCACCTTACAGAGGACCAGTTAGGCTGTGC...	6:<AA=
12	30PPJAAXX090125:1:43:1756:518#0	163	chr1	34825	0	76M	=	34871	-121	CATCACTATTTGTCAGGGAAGTACTAATTAAAACAAAAATGAGATGTCCCCTCAAACCTGTTAG...	B>3?CB

Fig. 4 Tabular display of a BAM file

Future work

SPAI is an on-going project and under active development. In future release, we plan to move the indel calling steps of the tools to a cloud based service which will completely reduce the computational burden of user's computer. Moreover, the future release will not require the user to download large alignment files. Instead, user will supply URL to the alignment file not only from the FTP server of the 1000 Genomes project but also from other sources and SPAI will fetch the file and produce result in the cloud based service. We will also let the user to open account in the cloud based service and the user's previously obtained results will be stored in the account so that these results can be used later instead of running the tools again. We plan to keep including newly developed indel calling tools and the next release will include some highly used tools such as GATK Haplotype Caller, Platypus, FreeBayes, and indelMINER. Moreover, we will also add utility tools such as UPS-indel - a tool to find ambiguous indels [46]. In addition to that we also plan to include tools that display the effect of a list of indels in coding and non-coding regions. Batch processing feature will also be added which will allow the user to perform multiple tasks simultaneously.

Results and Discussion

In this paper we addressed two research questions that are given below:

What is the necessity of creating a GUI for existing indel calling tools?

Researchers heavily depend on existing command line programs for calling indels from their dataset as well as for downstream analysis to the problems of their research domain. Since most of these tools are very popular and widely used by the research community, the question which may arise is why don't we just keep these tools as they are right now? From our experience of working on indel projects, we realized that to explore different features of these existing tools by changing parameters and/or by changing inputs, users need to write the command every time. This needs some expertise of command line programming. Moreover exploring tools by writing command every time causes lack of usability of these tools. To solve this usability problem and to overcome the requirement of command line programming expertise, we designed SPAI, a Graphical User Interface (GUI) based tool. Being a GUI based tool, SPAI allows users to explore the features of existing indel calling tools just by selecting input through a regular file browser and setting parameters by writing it in a text box. This not only saves time required for writing commands, but also gives users a better user experience. Moreover, since SPAI includes multiple indel calling tools in the same platform, user can run multiple tools at the same time for same inputs without writing a single line of command.

How SPAI can help in downstream analysis of indels?

After calling indels, the next step is the downstream analysis of the called result. The research question that we addressed is how SPAI can help in this context? SPAI comes with a list of known indels [42] for human genome which has been used as benchmark dataset by many researchers. After importing indel calling results from different tools, SPAI compares those indels with the above mentioned benchmark dataset. SPAI produces the comparison results in graphical format and also provides statistics such as recall, precision, and F-measure. This feature of SPAI allows users to assess the performance of the indel calling tools based on these matrices. Moreover, user can also supply their own benchmark dataset and list of indels produced by their own tool. Since SPAI produces graphs and charts with performance comparison matrices, user can easily assess the performance of their tools without doing these comparisons by themselves. This also saves time and ensures better user experience.

Conclusions

Indels constitute the most common form of structural variation in human genome and have been found to be responsible in causing diseases by abolishing gene functions. In addition to that, indels can influence human traits and gene expression and therefore can be used as genetic marker. All these statements lead to the necessity of variant profiling which should be achievable as whole genome sequencing at individual level is now possible because of Next Generation Sequencing (NGS) technology. A good number of indel calling tools have been developed that can be used for variant profiling purpose. However, all of these are command line based which require certain expertise from Computer Science (CS). As evolutionary genomics is a multi-disciplinary research area, people from non-CS background are also involved in indel calling research and should be able to use these tools without prior knowledge of command line programming. Here we introduce SPAI (Single Platform for Analyzing Indels) which provides user with an interactive platform to use popular indel calling tools using a user friendly GUI and perform different analyses without knowing any command line programming. We believe that people especially from non-CS background will find SPAI really useful while performing their indel calling research.

Acknowledgements
The work is partially supported by an NSF grant OCI-1124123 to L. Zhang.

Declarations
This article has been published as part of BMC Genomics Volume 17 Supplement 5, 2016. Selected articles from the 11th International Symposium on Bioinformatics Research and Applications (ISBRA '15): genomics. The full contents of the supplement are available online https://bmcgenomics.biomedcentral.com/articles/supplements/volume-17-supplement-5.

Funding
Publication of this article was funded by the Virginia Tech's Open Access Subvention Fund.

Availability of data and materials
Input to SPAI is the BAM file and reference genomes which are publicly available at the FTP server of the 1000 Genomes project. BAM file for any individual can be collected from this URL: ftp://ftp-trace.ncbi.nlm.nih.gov/1000genomes/ftp/data/. The reference genome sequence can be found from this URL: ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/.

Authors' contributions
MSH conceived the idea and performed the coding and computational experiments. LZ participated in the design of the tool and supervised the project. MSH and LZ wrote the paper. Both authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Published: 31 August 2016

References

1. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet. 2010;19(R2):R131–6.
2. Emde A-K, Schulz MH, Weese D, Sun R, Vingron M, Kalscheuer VM, et al. Detecting genomic indel variants with exact breakpoints in single-and paired-end sequencing data using SplazerS. Bioinformatics. 2012;28(5):619–27.
3. Ball EV, Stenson PD, Abeysinghe SS, Krawczak M, Cooper DN, Chuzhanova NA. Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat. 2005;26(3):205–13.
4. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas N, et al. The human gene mutation database: 2008 update. Genome Med. 2009;1(1):13.
5. Collins FS, Drumm ML, Cole JL, Lockwood WK, Woude GV, Iannuzzi MC. Construction of a general human chromosome jumping library, with application to cystic fibrosis. Science. 1987;235(4792):1046–9.
6. Warren ST, Zhang F, Licameli GR, Peters JF. The fragile X site in somatic cell hybrids: an approach for molecular cloning of fragile sites. Science. 1987;237(4813):420–3.
7. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. Hum Mol Genet. 2010;19(R2):R125–30.
8. Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature. 1988;332:164–6. doi:10.1038/332164a0.
9. Ostertag EM, Kazazian HH. Retrotransposition and Human Disorders. eLS. 2006. doi:10.1038/npg.els.0005492. Available from http://onlinelibrary.wiley.com/doi/10.1038/npg.els.0005492/full.
10. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH. SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet. 2003;73(6):1444–51.
11. Ostertag EM, Kazazian Jr HH. Biology of mammalian L1 retrotransposons. Annu Rev Genet. 2001;35(1):501–38.
12. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, et al. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res. 2006;16(9):1182–90.
13. Cheung VG, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat Rev Genet. 2009;10(9):595–604.
14. Väli Ü, Brandström M, Johansson M, Ellegren H. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. BMC Genet. 2008;9(1):8.
15. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol. 2007;5(10):e254.
16. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. nature. 2008;452(7189):872–6.
17. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, et al. The diploid genome sequence of an Asian individual. Nature. 2008;456(7218):60–5.
18. Kim JI, Ju YS, Park H, Kim S, Lee S, Yi JH, et al. A highly annotated whole-genome sequence of a Korean individual. Nature. 2009;460(7258):1011–5.
19. Ahn S-M, Kim T-H, Lee S, Kim D, Ghang H, Kim D-S, et al. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. Genome Res. 2009;19(9):1622–9.
20. Shigemizu D, Fujimoto A, Akiyama S, Abe T, Nakano K, Boroevich KA, et al. A practical method to detect SNVs and indels from whole genome and exome sequencing data. Scientific Reports. 2013;3:2161. doi:10.1038/srep02161. Available from http://www.nature.com/articles/srep02161?WT.ec_id=SREP-631-20130801.
21. Hasan MS, Zhang L. P-Dindel: A multi-thread based tool for calling indels from short reads. In: Short Abstracts of the 11th International Symposium on Bioinformatics Research and Applications. 2015. Norfolk, VA. 71-4. Available from http://www.cs.gsu.edu/isbra15/sites/default/files/ISBRA12ShortAbstractsFinal.pdf.
22. Hasan MS, Wu XW, Zhang LQ. Performance evaluation of indel calling tools using real short-read data. Human Genomics 2015;9. doi:ARTN 20 10.1186/s40246-015-0042-2
23. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.
24. GATK HaplotypeCaller. www.broadinstitute.org/gatk/guide/article?id=4148. Accessed June 30, 2015.
25. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25(17):2283–5.
26. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568–76.
27. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25(21):2865–71.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
29. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. Dindel: accurate indel calls from short-read data. Genome Res. 2011;21(6):961–73.
30. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SR, Wilkie AO, et al. Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014;46(8):912–8.
31. Grimm D, Hagmann J, Koenig D, Weigel D, Borgwardt K. Accurate indel prediction using paired-end short reads. BMC Genomics. 2013;14(1):132.
32. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011;21(6):936–9.
33. Korbel JO, Abyzov A, Mu XJ, Carriero N, Cayting P, Zhang Z, et al. PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. Genome Biol. 2009;10(2):R23.
34. Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurles ME, et al. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res. 2010;20(5):623–35.
35. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods. 2009;6(9):677–81.
36. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. 2012. p. 1–9. Available from http://arxiv.org/pdf/1207.3907.pdf.

37. Ratan A, Olson TL, Loughran TP, Miller W. Identification of indels in next-generation sequencing data. BMC Bioinformatics. 2015;16(1):42.
38. Hasan MS, Zhang L, editors. SPAI: Single Platform for Analyzing Indels. In: Short Abstracts of the 11th International Symposium on Bioinformatics Research and Applications. 2015. Norfolk, VA. 75-8. Available from http://www.cs.gsu.edu/isbra15/sites/default/files/ISBRA12ShortAbstractsFinal.pdf.
39. Galitz WO. The essential guide to user interface design: an introduction to GUI design principles and techniques. 3rd ed. Indianapolis, IN: Wiley Pub; 2007. p. 3–10.
40. Hasan MS, Liu Q, Wang H, Fazekas J, Chen B, Che D. GIST: Genomic island suite of tools for predicting genomic islands in genomic sequences. Bioinformation. 2012;8(4):203–5.
41. Consortium GP. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56–65.
42. Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, et al. Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res. 2011;21(6):830–9.
43. Whelan C. Detecting and Analyzing Genomic Structural Variation Using Distributed Computing. 2014. 2-156. Available from http://digitalcommons.ohsu.edu/cgi/viewcontent.cgi?article=7928&context=etd.
44. Whelan CW, Tyner J, L'Abbate A, Storlazzi CT, Carbone L, Sönmez K. Cloudbreak: accurate and scalable genomic structural variation detection in the cloud with MapReduce. arXiv preprint arXiv:13072331. 2013. p. 1–44. Available from http://arxiv.org/pdf/1307.2331v2.pdf.
45. BAMSeek. https://code.google.com/p/bamseek/. Accessed July 1, 2015.
46. Hasan MS, Li Z, Zhang L. UPS-indel: Universal Positioning System for Indels. 2016; Under Submission. Available from https://sourceforge.net/projects/ups-indel/.