

## Report on the ALLBIO minisymposium and workshop: “Next Generation Sequencing (NGS) methods for identification of mutations and large structural variants”



Laurent Falquet<sup>1</sup>✉, Grégoire Rossier<sup>2</sup>, Tiffanie Yael Maoz<sup>3</sup>

<sup>1</sup>University of Fribourg and Swiss Institute of Bioinformatics, Biochemistry Unit, Fribourg, Switzerland

<sup>2</sup>Vital-IT, Training & Outreach, Swiss Institute of Bioinformatics, Lausanne, Switzerland

<sup>3</sup>Weizmann Institute of Science, Lab of Prof. Avi Levy, Rehovot, Israel

Received 10 April 2014; Published 14 April 2014

Falquet L *et al.* (2014) *EMBnet.journal* 20, e766. <http://dx.doi.org/10.14806/ej.20.0.766>

This workshop was organised as one of the validation training workshops of the AllBio FP7 Coordination Action. The [AllBio project](#)<sup>1</sup> aims to transfer human genome-oriented bioinformatics methods to non-model organisms.

Following a first round of test-case proposals from all over Europe, a selection of 15 test-cases was presented and discussed in detail during an initial [workshop in December 2012 in Milano](#)<sup>2</sup>, Italy, organised by Dr. Andreas Gisel, CNR Institute for Biomedical Technologies of Bari (IT). From this event, seven test-cases were selected for ‘hackathon’ sessions, where real data were analysed jointly by computer scientists, bioinformaticians and biologists.

Our test-case for identification of large structural variants (insertions, deletions, inversion, translocations, *etc.*...) immediately attracted a lot of interest, given the current difficulty in predicting large structural variants (SVs) in all organisms, but in particular in polyploid genomes, as in plants, and chimeric genomes, as in cancer. The first hackathon session was planned during the pre-

paratory phase, using regular video meetings to define the goals and discuss ideas on how to solve them. It was decided that a Virtual Machine (VM) would be set up to provide a standardised platform as a benchmark method to evaluate and compare all tools. A small group of participants, led by Dr. Yael Maoz (Weizmann Institute, IL), met in March 2013 in Amsterdam (NL) with the support of the local organiser, Prof. Gert Vriend, and the computational support of [SURFsara](#)<sup>3</sup>, for an intensive hackathon session where one biologist, three bioinformaticians and four computer scientists met to solve a real case. The outcome was a preliminary version of a VM, hosting many tools and a benchmark data-set. During the following months, regular video meetings allowed the participants to combine their expertise in order to improve this first version of the VM.

In August 2013, a second hackathon session was hosted in Nijmegen (NL) and allowed testing of a merging tool that combines results of individual structural-variation prediction tools in a hierarchical manner, aiming to reduce false-positive calls. The results being promising, a publication and a validation workshop were planned for March 2014, in Lausanne, Switzerland.

The workshop was expanded to include a mini-symposium on the first day, with eight invited speakers and more than 75 participants from all over Europe (Figure 1). After the usual welcome address by the local organisers, Yael Maoz detailed the work of the hackathon team and the proposed outcomes of the AllBio project:

- creating an automated standardised pipeline for testing new tools and/or for testing existing tools on non-model organisms;
- identifying the best tool(s) for SV prediction through benchmarking;
- providing a statistically sound method of merging SV calls;
- a tool called SV-Autopilot: Structural Variation AUTomated Pipeline Optimisation Tool (submitted for publication).

Alexandre Reymond gave an amazing talk on the role of large SVs in human, showing the balancing effect of deletion and duplication of the same genomic location on diseases such as autism and schizophrenia, and their links with obesity (Zufferey *et al.*, 2012).

<sup>1</sup> [www.allbioinformatics.eu/doku.php](http://www.allbioinformatics.eu/doku.php)

<sup>2</sup> [www.allbioinformatics.eu/doku.php?id=public:bioinf](http://www.allbioinformatics.eu/doku.php?id=public:bioinf)

<sup>3</sup> <https://www.surfsara.nl/>



Figure 1. Participants at the mini-symposium.

Tobias Rausch presented the Delly software tool (Rausch *et al.*, 2012), and the various uses in many large-scale population genomics analyses. He also mentioned that inversions are usually not found alone: in more than 60% of cases, inversions are associated with deletions or duplications.

Bart Deplancke gave a brilliant talk on the analysis of large genomic variants in *Drosophila* lines identified using PrInSeS (Massouras *et al.*, 2010), and their effects (Massouras *et al.*, 2012).

Tobias Marshall gave a detailed talk on MATE-CLEVER, an extension of CLEVER (Marschall *et al.*, 2013) used to discover 'twilight zone' variants and their genotypes.

Valentina Boeva showed various tools to discover large SVs in cancer cell lines (Boeva *et al.*, 2013).

Yogesh Paudel introduced copy-number variation methodology used to analyse domestication events in pig lines (Paudel *et al.*, 2013).

Can Alkan provided a detailed account of the characterisation of mobile element insertions in humans and apes (Hormozdiari *et al.*, 2013).

The hands-on workshop on the second day joined 30 participants and six speakers. The attendance would have been larger, but was limited to this number for practical purposes.

Participants were asked to download and install a pre-configured Ubuntu Virtual Machine using VirtualBox software on their laptops.

Yael Maoz explained the concepts of the project and of the hackathon sessions, while building and use of the VM was detailed by Wai Yi Leung (Leiden University Medical Center). Participants were able to test the VM on a subset of the original data only, as the whole genome analysis would run for too long on a laptop. The results were then detailed and discussed by Tobias Marschall and Yael Maoz. Visualisation of the results with the Integrative Genomic Viewer (IGV) was presented by Laurent Falquet and Yael Maoz (Figure 2).

In the afternoon, the participants were divided into two groups: those wishing to analyse their own data, and those wishing to discuss issues and solutions for detecting large SVs in planned projects. The workshop ended with a summary given by Yael Maoz.

In conclusion, this workshop validates the outcome of our test-case dealing with large SVs. We have shown that a small group of volunteers working on a part-time basis can develop new methods and tools. We believe that the SV-AUTOPILOT VM that we developed will be very useful both for biologists looking to get the best variant predic-





Figure 2. Hands-on workshop. In this picture, Laurent Falquet illustrates IGV.

tions, and for bioinformaticians seeking to evaluate their software performance against existing tools.

### Acknowledgements

This work was supported by the AIBio FP7 work programme KBBE-2011-5-289452 «FOOD, AGRICULTURE AND FISHERIES, AND BIOTECHNOLOGY».

We thank the COST Action, SeqAhead, for supporting travel grants of participants and providing local organiser support, and the Vital-IT group of the SIB Swiss Institute of Bioinformatics, for local organisation.

### References

- Boeva V, Jouannet S, Daveau R, Combaret V, Pierre-Eugène C et al. (2013) Breakpoint features of genomic rearrangements in neuroblastoma with unbalanced translocations and chromothripsis. *PLoS One* **8**, e72182. <http://dx.doi.org/10.1371/journal.pone.0072182>
- Hormozdiari F, Konkel MK, Prado-Martinez J, Chiatante G, Herraiz IH et al. (2013) Rates and patterns of great ape retrotransposition. *Proc Natl Acad Sci U S A* **110**, 13457–13462. <http://dx.doi.org/10.1073/pnas.1310914110>
- Marschall T, Hajirasouliha I, Schönhuth A (2013) MATE-CLEVER: Mendelian-inheritance-aware discovery and genotyping of midsize and long indels. *Bioinformatics* **29**, 3143–3150. <http://dx.doi.org/10.1093/bioinformatics/btt556>
- Massouras A, Hens K, Gubelmann C, Uplekar S, Decouttere F et al. (2010) Primer-initiated sequence synthesis to detect and assemble structural variants. *Nat Methods* **7**, 485–486. <http://dx.doi.org/10.1038/nmeth.f.308>
- Massouras A, Waszak SM, Albarca-Aguilera M, Hens K, Holcombe W et al. (2012) Genomic variation and its impact on gene expression in *Drosophila melanogaster*. *PLoS Genet* **8**, e1003055. <http://dx.doi.org/10.1371/journal.pgen.1003055>
- Paudel Y, Madsen O, Megens H-J, Frantz LAF, Bosse M et al. (2013) Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics* **14**, 449. <http://dx.doi.org/10.1186/1471-2164-14-449>
- Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V et al. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339. <http://dx.doi.org/10.1093/bioinformatics/bts378>
- Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM et al. (2012) A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J Med Genet* **49**, 660–668. <http://dx.doi.org/10.1136/jmedgenet-2012-101203>