

## Power and limits of capture-based, targeted DNA resequencing for mutation detection

**Fabrice Lopez, H  l  ne Holota, Fran  ois-Xavier Th  odule and Jean Imbert**

TAGC UMR \_ S 928, Inserm, Universit   de la M  diterran  e, Marseille, France

<http://www.yourwebsite.org/>

The IBISA TGML platform (Sci.Dir.: Dr. J. Imbert) is integrated to TAGC (Inserm U928, Dir.: Dr C. Nguyen) located on the Science Park of Luminy (Universit   de la M  diterran  e). It offers access for academics and companies to transcriptomic and functional genomic studies. The TGML platform and TAGC provide expertise in the analysis of various types of DNA microarrays and sequencing dataset. Our researchers and engineers actively contribute to the development of new computational tools and data processing pipelines. The TGML platform is member of the France-G  nomique network. The high throughput sequencing service (TGML DeepSeq) is equipped with a LifeTech SOLiD4 sequencer that can produce up to 100 Gb (fragments 50 nt). Barcoding allows sequencing up to 256 samples in one run. Upgrade to SOLiD 5500XL (300 Gb, 75 nt) is scheduled Fall 2011 as well as the purchase of an Ion Torrent PGM machine for a fast and cost-effective sequencing alternative for smaller sized projects. Applications performed as a service for external users or in collaboration include: full exome and targeted DNA resequencing (*Homo sapiens*, *Mus musculus*, etc.) with customized capture design on microarrays, ChIP-seq, FAIRE-seq, Mnase-seq (Epigenomics), and some projects of full resequencing for small genomes. Collaborators and clients include teams from CIML, IBDML, CRCM, IAB in Grenoble, a partnership with the GIS Institut GIS Maladies Rares (Paris, Marseille, Dijon, Montpellier, etc.), etc. We are planning to implement shortly: whole transcriptome analysis (lncRNA), SAGE-seq, DNaseI-HS-seq, de novo bacteria sequencing, FAIRE-seq, etc. During the last 2 years we have acquired a robust experience in targeted genomic resequencing and we are currently developing of a new bioinformatics pipeline for the characterization of genomic variants (SNPs and small InDels) and a new Java-based graphic software named GeVarA for Genomic Variant Analyzer.

I will present the power and the limit of these approaches with an emphasis on the challenges faced by the bioinformatician and by our computing and data storage resources, as well as our ongoing solutions.

### References

1. Lopez,F., Textoris,J., Bergon,A., Didier,G., Remy,E., Granjeaud,S., Imbert,J., Nguyen,C., and Puthier,D. (2008). TranscriptomeBrowser: a powerful and flexible toolbox to explore productively the transcriptional landscape of the Gene Expression Omnibus database. *PLoS ONE* 3, e4001.
2. Benoukraf,T., Cauchy,P., Fenouil,R., Jeanniard,A., Koch,F., Jaeger,S., Thieffry,D., Imbert,J., Andrau,J.C., Spicuglia,S., and Ferrier,P. (2009). CoCAS: a ChIP-on-chip analysis suite. *Bioinformatics* 25, 954-955.
3. Pekowska,A., Benoukraf,T., Zacarias-Cabeza,J., Belhocine,M., Koch,F., Holota,H., Imbert,J., Andrau,J.C., Ferrier,P. and Spicuglia,S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J.* doi:10.1038/emboj.2011.295.

### Relevant Web sites

4. <http://tagc.univ-mrs.fr/welcome/spip.php?rubrique1>
5. <http://tagc.univ-mrs.fr/welcome/spip.php?rubrique2>