

Efficient Annotation of Variant Calls: Next Generation Sequencing

Muntaser Ibrahim* and Mahmoud Musa

Institute of Endemic Diseases, University of Khartoum, Medical Campus, Khartoum, Sudan

Introduction

The tremendous and unprecedented insights provided by next generation sequencing into genome functions, variations and interactions promises an enormous shift in our attitude towards individual and population genetics, both in health and disease. Concepts and paradigms could be verified or nullified based on large complete sets of data rather than few genes or inference from large sets of comparison with substantial gaps like genome wide association studies (GWAS). Functional variant(s) may be associated with phenotypes at a “personalized” level. Rare variants underlying common diseases-and even underlying variations in health-are a frequent encounter in the genomic era. Simple inductive generalizations would hastily blur the lines between what is a true abnormality and what is not.

Effective size, imputation and generalizations: The caveat of doing genomics against a background of large effective size with lack of representative sampling [1] is the problem of establishing a reliable reference. The extent of existing variation is unknown or quite large at best assumptions. Experiences gained especially from GWAS have shown the pronounced impact of population structure. Imputation will only compensate for a certain threshold of frequent variants within the population. This has its implications on complex disease mapping [2]. We have witnessed a huge shift in imputation accuracy between the Hap Map and 1000 Genome [3]. But can we ever be satisfied? Dealing with exome and genome sequencing in Africa, we find ourselves perplexed. The evolutionary history of many African populations is deeply rooted making imputation a risky practice.

Efficient Annotation of Variant Calls

When making variant calls and subsequent annotations, the “reliable” reference should take into account these issues of structure and large effective size. The premise of a putative ancestral genome that may be reconstructed from existing genomes may be flawed. Here we mention two arguments that support this: firstly, human populations in and out of Africa were influenced and shaped by drift and serial founder effects, and secondly, functional variants are multiple, formed through an extended evolutionary history in Africa and retained within its extant populations. Both assumptions have multiple backings and could be tested afresh in global genomic data and especially African genomes. Eventually and with such scenarios in place, genomics is at a greater challenge. Identifying association and causality for any disease model departing from the simple form of “common variant common phenotype” would be a big hurdle to overcome. Studying an exome for example, we are faced by thousands of variants; many of them are peculiar to the background population which eliminates the luxury of blindly utilizing available allele frequency data as a filter for variants. Looking for changes in “common” disease genes might not hold the answer either keeping in mind what we argued on the evolutionary history of these populations. Other genes might take their place, promoting other models, for example the rare variant model. Using familial or population controls is one way around. Africa is bound to face even greater another challenge for personalized medicine specially with the practice of projecting annotation based on published data, for example from the 1000 genome, and other relevant databases without accompanying the structure and genetic history into the analysis.

Effective Genomics

Extensive efforts should be made in order to identify not only the gene functions but the unique epistatic effects and interaction networks. Using systems biology and network analysis in studying phenotypes as a substitute or a complementary approach to single gene prioritization in African populations will yield valuable information. Networks are regarded as conserved evolutionary units that can undergo selection [4,5]. Evolutionary plasticity of genetic interaction networks [6] adds another interesting perspective. Genes that are conserved between species have been functionally maintained since they last shared a common ancestor. On the other hand, non-additive genetic interactions are at best poorly conserved, as they are the direct consequence of network rearrangements that are themselves not conserved. The notion that networks should be conserved at a functional level is plausible given the fluidity of cellular processes and gene regulation on a background of differing variations. Hence, loss of same or different gene sets can affect different or certain phenotypes, respectively, in different populations, the mere link being the orchestrated interaction networks affected by their demise. We take the example of TP53, a frequent finding in cancer in Western populations but not necessarily in African populations. Other genes with interaction centrality, for instance HuR/ELAVL1, might underly oncogenesis instead. We have recently touched upon the centrality of ELAVL1 and it's possible role in conjunction with other viral oncogenesis in African populations [7]. This by no means is a call to throw all what we have. Existing genotype and phenotype databases are certainly the basis for annotation. Yet, another level of scrutiny is warranted. The formulation of standards for this level should be our next priority.

In summary, effective genomic analysis in founder populations should take into account the complex evolutionary history. Available annotations should be interpreted with caution. The large effective population size merits the utilization of population specific data and the use of systems approaches to prioritize variations.

References

1. Elhassan N, Gebremeskel EI, Elnour MA, Isabirye D, Okello J, et al. (2014) The Episode of Genetic Drift Defining the Migration of Humans out of Africa Is Derived from a Large East African Population Size. *PLoS One* 9: e97674.
2. Campbell MC, Tishkoff SA (2008) African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual review of genomics and human genetics* 9: 403-433.

*Corresponding author: Muntaser Ibrahim, Institute of Endemic Diseases, University of Khartoum, Medical Campus, Khartoum, Sudan, Tel: +249 11 310102; E-mail: mibrahim@iend.org

Received December 10, 2021; Accepted December 28, 2021; Published January 31, 2021.

Citation: Ibrahim M, Musa M (2021) Efficient Annotation of Variant Calls: Next Generation Sequencing. *Next General Sequenc & Applic* 7:1; 008. doi:10.4172/2469-9853.S1-008

Copyright: © 2016 Ibrahim M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

-
3. J. Wang L, Rankinen T, Bouchard C, Rao DC (2012) Performance of Genotype Imputations Using Data from the 1000 Genomes Project. *Human Heredity* 73: 18-25.
4. Luisi P, David A, Marc P, Fares MA, Jaume B, et al. (2015) Recent Positive Selection Has Acted on Genes Encoding Proteins with More Interactions within the Whole Human Interactome. *Genome Biology and Evolution* 7: 1141-1154.
5. Koch EN, Costanzo M, Bellay J, Deshpande R, Chatfield-Reed K, et al. (2012) Conserved rules govern genetic interaction degree across species. *Genome Biology* 13: R57.
6. Tischler J, Lehner B, Fraser AG (2008) Evolutionary plasticity of genetic interaction networks. *Nature Genetics* 40: 390-391.
7. Suleiman H, Koko ME, Nasir WH, Elfateh O, Elgizouli UK, et al. (2015) Exome sequencing of a colorectal cancer family reveals shared mutation pattern and predisposition circuitry along tumor pathways. *Front Genet* 6: 288.

This article was originally published in a Volume 7:issue1 handled by Editor(s). Dr. Jianping Wang, University of Florida, USA.