LETTER

The future of the human SNP identification: Which individuals to sequence?

We applaud Ionita-Laza et al.'s important contribution on the number of "unseen" SNPs and the number of individuals to be sequenced sensibly to identify all human SNPs within reason (1). We would like to note a matter of some debate in the interested scientific community: namely, whether random individuals should be sequenced, or those with extreme phenotypes (ref. 2 and www.1000genomes.org/). This is not simply an esoteric academic question but one that bedevils most sequencing projects. We firmly believe that biological and medical considerations are important, if not critical, in identifying high-risk individuals that are very likely to bear genetic variants that may predispose them to a particular disease, whereas, at the opposite end of the spectrum, low-risk individuals may bear alleles that may protect them from disease, etc. We propose that such individuals with "extreme phenotypes" should be sequenced preferentially to identify phenotypically and/or medically relevant allelic variants. Therefore, various risks and phenotypes may have to be taken into account when modelling the SNP analyses of Ionita-Laza et al. We are certain that such a biologically and medically informed approach would reduce the number of individuals needed to be sequenced based on their risks or protection, leading to an increased chance of identifying important variants.

In summary, we applaud the illuminating efforts of Ionita-Laza and colleagues (1) and would recommend that presequencing analyses take into account susceptibility to disease and protection from disease to select those highly informative individuals for sequencing whom are most likely to bear significant genetic variants.

***Juergen K. V. Reichardt*[a,1] *and Ruty Mehrian-Shai*[b]**

[a]*University of Sydney, Medical Foundation Building (K25), Plunkett Chair of Molecular Biology (Medicine), Camperdown NSW 2006, Australia; and* [b]*Institute for Genetic Medicine, University of Southern California, Keck School of Medicine, Los Angeles, CA 90089*

1. Ionita-Laza I, Lange C, Laird NM (2009) Estimating the number of unseen variants in the human genome. Proc Natl Acad Sci USA 106:5008–5013.
2. Nebert DW (2000) Extreme discordant phenotype methodology: An intuitive approach to clinical pharmacogenetics. *Eur J Pharmacol* 410:107–120.

Author contributions: J.K.V.R. analyzed data; and J.K.V.R. and R.M.-S. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: jreichardt@med.usyd.edu.au.

www.pnas.org/cgi/doi/10.1073/pnas.0903681106