Successful precision medicine will require more accurate genome sequencing

Large areas of medically important genes fall within troublesome regions of the human genome, where it is currently difficult to obtain accurate sequence information, according to research published in the open access journal [*Genome Medicine*](http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0269-0). On average, one fifth of each of these medically important genes is challenging for today’s gene sequencing methods to decipher, and the information in these gene regions may be key to a patient’s diagnosis or treatment plan.

To optimize medical care, an accurate account of each patient’s genetic code is needed to predict risk for disease and to select appropriate medication. The study by researchers from Stanford University highlights the medical consequences of sequencing errors.

Such errors include false positives (identifying genetic mutations that aren’t really there) as well as false negatives (failing to detect legitimate disease-causing mutations). Both can have profound consequences for patient care. For example, a false positive mutation in BRCA2, a well-known gene associated with hereditary breast and ovarian cancer, could lead to risk-reducing surgeries, such as double mastectomy and oophorectomy. Thus, a wrongly identified mutation could potentially lead to radical and unnecessary surgeries.

The Stanford team used a gold-standard genome sequence, provided by the US National Institute of Standards and Technology (NIST). This genome, belonging to a female of European ancestry, had been previously sequenced with five different sequencing technologies. The NIST team combined the results from all five technologies to develop a reliable consensus sequence in regions of the genome where the technologies agreed. A reliable consensus was achieved for just 77% of this donor’s genome.

Looking at how these “high confidence” areas of the donor’s genome overlap with 3,300 genes known to cause human disease, the researchers found that for 593 of these genes, less than half of the crucial protein-coding regions are in areas that can reliably be sequenced.

There is a group of 56 disease genes regarded as most medically “actionable” by the American College of Medical Genetics and Genomics (ACMG), including BRCA2. ACMG guidelines now require clinical genetic testing labs to screen all patients undergoing exome or genome sequencing for disease-causing mutations in these 56 genes, which are involved in treatable conditions ranging from hereditary cancer to life-threatening cardiac arrhythmias. A patient might initially undergo sequencing to identify the cause of their autism, for example, yet would also be informed of an incidental finding in BRCA2, with the goal of predicting or even preventing disease.

Yet for these medically-important genes, the Stanford researchers found that only 80% of each gene’s protein-coding regions, on average, can be sequenced with confidence.

This study also shows that the majority of disease-causing mutations identified to date fall within easy-to-sequence areas. Specifically, among disease-causing mutations catalogued in the database ClinVar, more than 80% fall within high-confidence regions of the NIST genome. Furthermore, the overwhelming majority of these ClinVar mutations (greater than 98%) are in stretches of unique DNA sequence, long known to be easier to sequence.

These findings highlight the need for sequencing methods that better penetrate hard-to-sequence regions of the genome, accurately revealing disease-causing mutations there that may currently be obscured.

Lead author Rachel Goldfeder, from Stanford University, says, “As this technology moves from the research lab to the clinic, we need to be able to accurately and reliably sequence entire genomes, because incorrect sequence information can lead to inappropriate medical care. The good news is that, in this case, 77% of the donor’s genome was reliably sequenced using current methods. The challenge now is to focus our efforts on the other 23%—namely, on regions of the genome that remain elusive. Only then can we realize the full potential of precision medicine.”

**-ENDS-**