CrossMark
click for updates

OPINION ARTICLE

REVISED "The molecule's the thing:" the promise of molecular modeling and dynamic simulations in aiding the prioritization and interpretation of genomic testing results [version 3; referees: 2 approved, 1 approved with reservations]

Gavin R. Oliver[1,2], Michael T. Zimmermann[1], Eric W. Klee[1,2], Raul A. Urrutia[2,3]

[1]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905, USA
[2]Center for Individualized Medicine, Mayo Clinic, Rochester, MN, 55905, USA
[3]Laboratory of Epigenetics and Chromatin Dynamics, Epigenomics Translational Program, Gastroenterology Research Unit, Departments of Biochemistry and Molecular Biology, Biophysics, and Medicine, Mayo Clinic, Rochester, MN, 55905, USA

Abstract

Clinical genomics is now a reality and lies at the heart of individualized medicine efforts. The success of these approaches is evidenced by the increasing volume of publications that report causal links between genomic variants and disease. In spite of early success, clinical genomics currently faces significant challenges in establishing the relevance of the majority of variants identified by next generation sequencing tests. Indeed, the majority of mutations identified are harbored by proteins whose functions remain elusive. Herein we describe the current scenario in genomic testing and in particular the burden of variants of uncertain significance (VUSs). We highlight a role for molecular modeling and molecular dynamic simulations as tools that can significantly increase the yield of information to aid in the evaluation of pathogenicity. Though the application of these methodologies to the interpretation of variants identified by genomic testing is not yet widespread, we predict that an increase in their use will significantly benefit the mission of clinical genomics for individualized medicine.

This article is included in the Rare diseases channel.

Open Peer Review

Referee Status: ☑☑❓

	Invited Referees	
	1	**2**	**3**

REVISED
version 3
published
25 Jul 2016

REVISED		
version 2 published 21 Jun 2016	☑ report	☑ report	
	↑	↑	
version 1 published 27 Apr 2016	❓ report	❓ report	❓ report

1   **Saverio Alberti**, "G. d'Annunzio" University of Chieti-Pescara Italy

2   **Attila Bérces**, Omixon, Inc. Hungary

3   **Anita Grigoriadis**, King's College London UK

Discuss this article

Comments (1)

Corresponding author: Gavin R. Oliver (oliver.gavin@mayo.edu)

Competing interests: No competing interests were disclosed.

In under a decade, sequencing of a human genome moved from a three billion dollar, multi institutional effort, to a common research assay. Now, 16 years after completion of the draft sequence, genomic assays are an increasingly prevalent component of clinical testing. Oncology, hematology and the diagnosis of rare, Mendelian genetic disorders (diagnostic odyssey) have particularly benefited from the increased genetic testing resolution afforded by modern sequencing technologies. The commercialization of clinical assays to profile patients' somatic or germline genomes is creating the potential for higher resolution diagnoses and individualized treatment options. Stories of the resounding success of such efforts have been widely and justifiably publicized, and have now captured imaginations well beyond the laboratory or clinic; even so far as the White House and US Congress. The President's Precision Medicine Initiative[1] has now been signed into existence and guarantees expanded exploration within this burgeoning area of health sciences research.

While sequencing technologies enable the interrogation of entire genomic, transcriptomic and epigenomic repertoires, current clinical implementations remain heavily biased toward the protein coding component of the DNA, either in the form of whole-exome sequencing or gene panel testing. This is a reflection of both the relative maturity of the field of molecular genetics and the clinical familiarity with both it and the concept of molecularly targeted therapy, which currently represents the major paradigm of tailored cancer treatment. Limiting clinical interrogation of a patient genome to a coding subset may seem counterintuitive in the era of whole-genome sequencing, but the reality is that our understanding of non-coding genomic variation and its role in disease remains in its infancy, particularly with regard to clinical actionability. The explosion of genomics research has increased our understanding of the human genome and the wide-ranging functions it encodes, however most regions of the genome remain uncharacterized and poorly understood. Even with genomic testing limited to protein coding regions, the majority of sequence variants detected go unreported or are classified as variants of uncertain significance (VUSs) due to their uncertain impact on normal physiological function.

For the researchers and clinicians on the ground of this nascent field, enthusiasm is high but it is tempered with frustration due to the as-yet high rate of cases for which genomic testing remains insufficiently informative (around 75% for diagnostic odyssey)[2,3].

A recurring scene is the convening of large, multidisciplinary teams of clinical and research staff to pore over lengthy lists of genomic sequence variants, exchange professional opinion and debate next steps. However, in the end, these involved efforts often fail to identify variants with highly confident causal or mechanistic relationships with the disease phenotype. In oncology, the knowledge gained for treatment selection is often compelling, but with little prior evidence, most are difficult to actualize. For example, identification of a novel missense mutation in the functional domain of a known, druggable oncogene might logically appear to be a therapeutic target, but no information may exist linking the mutation to drug efficacy. It is this disconnect between novel genotypes identified and their link to the patient's phenotypes which thwarts our ability to further improve clinical decision making. Thus, there is a significant need to overcome this critical challenge to expand the success of genomic testing in individualized medicine.

The most mature applications of clinical exome and panel testing are the detection of focal substitutions (missense, non-sense or splice site), insertions and deletions (INDELS), or copy number changes. While copy number gains or losses can intuitively be interpreted as potential gain-of-function or loss-of function events respectively, point mutations and INDELS are less easily characterized in terms of their effects on protein function. Wet-lab assays designed to ascertain functional relevance of specific mutations are a highly desired solution, but they remain cumbersome, time-consuming and cost prohibitive in the great majority of cases. They are therefore misaligned with the high throughput nature of modern genomics. Meanwhile, computational methods of predicting the pathogenicity of genomic sequence alterations exist and are both amenable to high-throughput predictions and widely applied in the field, to the extent that they form a component of the American College of Medical Genetics and Genomics (ACMG) guidelines for the clinical interpretation of sequence variants[4]. These algorithms utilize varied information including evolutionary conservation, genomic position or basic protein-level structural information to assign probabilistic scores or categorical predictions of pathogenicity. Their accuracy varies widely, with alternative tools often producing conflicting predictions for the same variant. Furthermore, the predictions often lack contextualization in terms of biological effect. Algorithmic categorization of a variant as pathogenic offers little insight into the nature of its phenotypic effect and less indication still of whether it is relevant clinically or how to act upon it. Many researchers continue to work to improve these methods, but the burden of VUS interpretation persists and the need for a means to address this burden and facilitate clinical interpretation is widely felt within the field.

The aforementioned computational tools largely leverage genome-centric information. While there is much proven value to these data, the biological reality comprises many additional layers of complexity. Mutations affect atomic-level biophysical changes that may alter the structure and biochemical function of the genome's protein products. While we detect variants in DNA, we typically interpret their effect by inferring or confirming their deleterious effect on encoded proteins (Figure 1). Even the effects of regulatory variants are typically interpreted as altering the probability of a protein to be expressed or spliced into a given functional form. Thus ideally, any method used to assess pathogenicity of genomic variants should possess the ability to look beyond genomic annotation to a functional context, at an atomic resolution. This increased
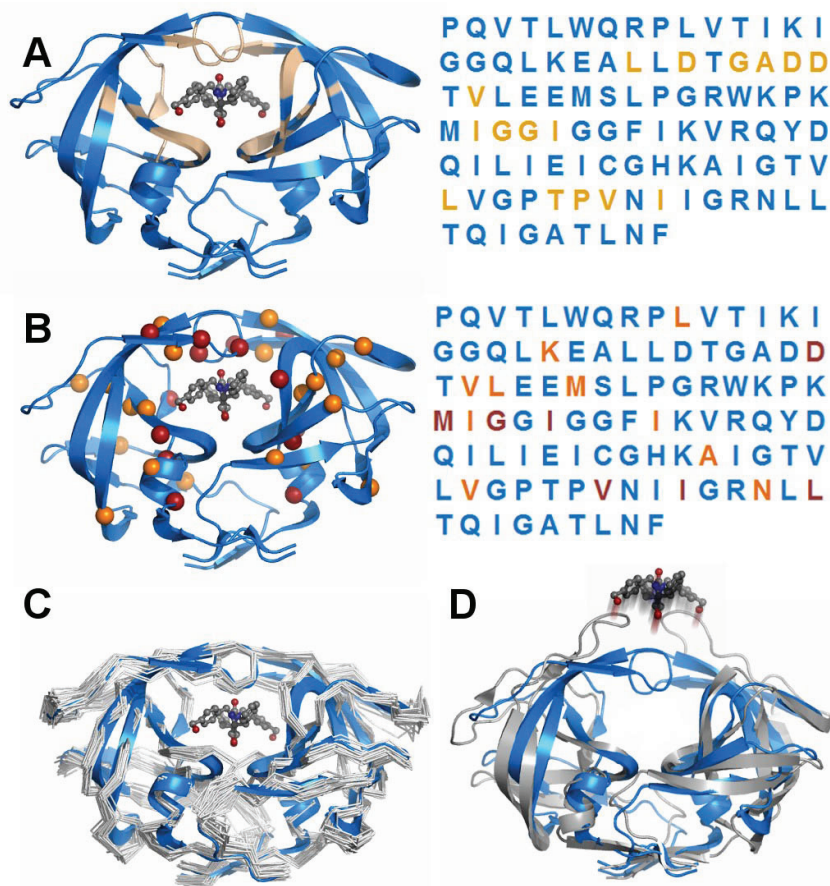
| P Q V T L W Q R P L V T I K I |
| G G Q L K E A L L D T G A D D |
| T V L E E M S L P G R W K P K |
| M I G G I G G F I K V R Q Y D |
| Q I L I E I C G H K A I G T V |
| L V G P T P V N I I G R N L L |
| T Q I G A T L N F |

Figure 1. The HIV-1 protease as a generic model system for computational biophysics[14–16]. HIV-1 protease has become a model system because of its disease relevance, the availability of mutational and drug binding data, and for its tractable size and molecular stability. The protein's function is to cleave HIV peptides into the functional proteins of the infectious HIV virion. This example illustrates generalized concepts of the applicability of protein modeling methods in clinical investigations A) *Ligand binding residues are spatially separated*. The functional protein dimer is shown with a pharmacologic inhibitor bound to the active site. Residues that are within 3.5Å of the inhibitor are highlighted in tan. The primary sequence is colored identically to the three-dimensional structure to indicate relative positioning of residues. It is apparent that the ligand binding portion of the protease consists of residues that are non-adjacent within the primary sequence, which illustrates an advantage of modelling over linear sequence analysis. B) *Drug resistance mutations tend to occur in residues within the active site*. Many mutations have been characterized that are associated with resistance to inhibitor drugs. While the sites of these mutations are also disjoint in sequence, nearly all of them fall into the same set of drug binding residues, indicating how modeling can enable prediction of a mutation's effect. C) *Structural effects of mutations beyond the active site*. Drug-resistance mutations in non-ligand-binding residues have been shown, using computational experiments, to impact the flexibility of the protein and therefore alter drug binding. Computational modeling has characterized the flexibility of the protein in multiple mutated states, illustrating the potential to predict the functional effect of mutations beyond an active site. D) *Dynamics of ligand-free protein*. Computational studies have the advantage of being able to simulate conditions that are difficult to assay experimentally, such as the dynamics of ligand-free forms in atomic detail.

resolution is likely to yield information that can add context to mutations, better identify the mechanism of pathogenicity, and combine with existing knowledge to facilitate in clinical decision making.

Computational biophysics and biochemistry aim to understand molecular function in a dynamic manner at an atomic resolution differing from their wet-lab counterparts methodologically but sharing the same goals. The most obvious advantages of computational approaches include the potential to test hypotheses *in silico* that would be difficult, costly, or intractable in the lab. Similar to laboratory experiments, computational calculations and simulations are most interpretable when they are well designed, test a

specific hypothesis, contain positive and negative controls, connect the data generated to the pre-existent knowledge in the field, and allow drawing further functional inferences. When all these conditions are met, the three dimensional and dynamic representation of the modeled mutations may add a significant value to the interpretation of a genomic test's finding.

Motivated by the conditions described, we have begun to apply molecular protein modeling and dynamic simulation techniques in the interpretation of genomic variants identified by next generation sequencing. These methodologies abandon the practice of regarding mutations as occurring in linear strings of nucleotides

or amino acids and instead offer a three dimensional, dynamic view, at an atomic resolution. Furthermore they are inherently compatible with the current state of clinical genomics testing and its predominant focus on point mutations within protein coding regions. The methods involve the computational generation, optimization and verification of a protein structural model, often based on homology to an experimentally determined protein structure[5,6]. *Ab initio* modeling is also possible, albeit with lower confidence. The model itself contains information about the linear amino acid sequence of the protein, along with the relative spatial coordinates of its atoms. Precise mathematical and biophysical parameters in the form of a force field are applied to the model to calculate energetic characteristics of the system. The methods are often mature and under reasonable conditions can be expected to produce a model with error comparable to that of a typical structure solved by nuclear magnetic resonance spectroscopy.

Molecular modeling allows us to visualize the manner in which proteins are folded to create a functional structure and to accurately simulate how this is disrupted by mutational events. Because we can visualize atomic bonds, mutations which disrupt *inter*-molecular interactions - for example between an enzyme and its substrate - can also be modeled. To illustrate their function, many enzymes are analogized to traditional tools. One example likens proteases and scissors; if a variant inhibits the closing motion of the scissors about their fulcrum, the blades cannot function; if a variant blocks entry of a material between the blades, then that material can no longer be cut. The reality, of course, is more complex. Proteins are flexible polymers that only achieve mechanistic accuracy by folding into complex and specific three dimensional conformations that restrain or focus thermal fluctuations towards collective motions that are typically part of the mechanism itself. Regions of the properly folded structure's surface that interact with other molecules, either for chemical modification (e.g. phosphorylation) or structural contacts (e.g. α/β tubulin assembly), can also be critical for function and modified by variants. In addition to the native shape of a protein, the ability of the linear amino acid polymer to achieve that shape is critical. Protein folding often occurs through progressive assembly of local structural elements or intermediates. If an intermediate is either stabilized or destabilized by a variant, the ability of the protein to achieve the native fold could be altered.

Of course, molecular modeling methods have limitations. The generation of a reliable model is often dependent on the pre-existence of experimentally determined homologous structures and even where these exist, they may provide only partial information[7]. A model is by nature an approximation of reality. Nonetheless, current techniques have achieved suitable accuracies such that they are frequently used[8-11] in applications including drug design, virtual screening, protein engineering and site-directed mutagenesis. With this in mind, their relatively slow uptake in the clinical setting is somewhat surprising. This fact may simply reflect the relative nascence of clinical genomics and the tendency of specialists to seek out fields in which their specialty is already known and accepted.

We encourage those working within or in proximity to the clinical genomics setting to engage in or promote increased exploration of these methods in their work. Our initial experiences of applying such techniques at the clinical-research boundary of genomics-driven oncology, hematology and diagnostic odyssey have been encouraging and have begun to inform decision-making. We are observing clear initial benefits in regard to variant prioritization, interpretation and validation, with several initial publications in preparation to highlight the value obtained. Of course, not everyone will possess the necessary scientific knowledge or technical skills to deploy these methods directly, but we propose that inter or intra-institutional collaborations may enable those lacking the requisite expertise to identify and access appropriate resources. Alternatively, several freely available online solutions exist[12,13] and enable a researcher or clinician to experiment with core modeling methods in the absence of extensive technical expertise. Such independent or collaborative exploration will open the door to deeper understanding of biological mutations and has the potential to inform clinical thinking. Wider use of these methods will enable the field to progress toward standardization of methodologies and approaches, and aid in the formulation of clinical best-practices that will serve to strengthen initiatives like the ACMG standards.

As we alluded to earlier in this article, wider 'omics' research is a broad and varied landscape of many alternative techniques that are continually revealing new findings with potential diagnostic and therapeutic relevance. Applications such as transcriptomics and epigenomics are beginning to develop clinical presence, but currently they lag behind DNA-based methods. Within the realm of DNA sequencing, the potential clinical relevance of non-coding variation is recognized and bioinformatics developments as well as major experimental initiatives like The Encyclopedia of DNA Elements (ENCODE) bring us closer to routine exploitation of such findings in the clinic. Multi-omic and systems-level research are also widening the scope in which we consider the molecular underpinnings of disease and these promise to have productive futures in clinical settings. Furthermore, novel treatments including oligonucleotide-based and immunogenic therapies are emerging, and move beyond the current prevailing targeted therapeutic paradigms. The case we have made for increased use of protein modeling focuses intentionally on the current state of clinical genomics testing and related diagnosis and treatment decision-making, rather than within these burgeoning research areas. Our opinion is that protein modeling satisfies a current and real need within clinical genomics and that its application should be immediately explored and encouraged. This is not to say it's clinical niche will be short-lived as many emerging techniques may well benefit from similar modeling methods being applied in either central or supportive roles, for example in nucleic acid structural modeling or in predicting neo-antigen affinities for MHC-class molecules.

In summary, genomic testing is assuming an increasingly prominent role in the clinical identification, prioritization, and interpretation of disease-associated genomic variants. While a few of these variants are known to be pathogenic, knowledge of the deleterious effects of the majority remain elusive. Laboratory methods remain gold-standards for functional characterization, but are generally incompatible with large-scale characterization of variant effects. Predictive algorithms are in some instances successfully applied to differentiate pathogenic variants from variants of uncertain significance, however these methods largely ignore measures

of protein structure, energies, molecular bonds, intermolecular interactions, post-translational modification effects, protein aggregation, and stability. Conversely, this information lies at the heart of molecular modeling and dynamic simulation, which collectively equip us more fully to grapple with interpreting the effects of VUSs. We strongly believe that these methodologies can be major components in forging the analytical pipeline for interpreting genomic testing in individualized medicine and advocate their increased deployment within variant characterization efforts.

Author contributions

EWK & GRO conceived the manuscript. GRO wrote the manuscript. MTZ generated figures and wrote the manuscript. RAU wrote the manuscript. All authors proof-read and approved the manuscript and have been involved in the modeling and interpretation of genomic variants in diagnostic odyssey, hematology and oncology studies within Mayo Clinic's Center for Individualized Medicine.

Competing interests

No competing interests were disclosed.

References

1. Collins FS, Varmus H: **A new initiative on precision medicine.** *N Engl J Med.* 2015; **372**(9): 793–795.
    PubMed Abstract | **Publisher Full Text**

2. Yang Y, Muzny DM, Xia F, *et al.*: **Molecular findings among patients referred for clinical whole-exome sequencing.** *JAMA.* 2014; **312**(18): 1870–1879.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Yang YP, Muzny DM, Reid JG, *et al.*: **Clinical whole-exome sequencing for the diagnosis of mendelian disorders.** *N Engl J Med.* 2013; **369**(16): 1502–1511.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Richards S, Aziz N, Bale S, *et al.*: **Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.** *Genet Med.* 2015; **17**(5): 405–424.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Martí-Renom MA, Stuart AC, Fiser A, *et al.*: **Comparative protein structure modeling of genes and genomes.** *Annu Rev Biophys Biomol Struct.* 2000; **29**: 291–325.
    PubMed Abstract | **Publisher Full Text**

6. Sali A: **Modeling mutations and homologous proteins.** *Curr Opin Biotechnol.* 1995; **6**(4): 437–451.
    PubMed Abstract | **Publisher Full Text**

7. Zhang Y: **Protein structure prediction: when is it useful?** *Curr Opin Struct Biol.* 2009; **19**(2): 145–155.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Ihle NT, Byers LA, Kim ES, *et al.*: **Effect of *KRAS* oncogene substitutions on protein behavior: implications for signaling and clinical outcome.** *J Natl Cancer Inst.* 2012; **104**(3): 228–239.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Duffy MR, Parker AL, Kalkman ER, *et al.*: **Identification of novel small molecule inhibitors of adenovirus gene transfer using a high throughput screening approach.** *J Control Release.* 2013; **170**(1): 132–40.
    PubMed Abstract | **Publisher Full Text**

10. Schmidt T, Bergner A, Schwede T: **Modelling three-dimensional protein structures for applications in drug design.** *Drug Discov Today.* 2014; **19**(7): 890–7.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Hillisch A, Pineda LF, Hilgenfeld R: **Utility of homology models in the drug discovery process.** *Drug Discov Today.* 2004; **9**(15): 659–669.
    PubMed Abstract | **Publisher Full Text**

12. Kelley LA, Mezulis S, Yates CM, *et al.*: **The Phyre2 web portal for protein modeling, prediction and analysis.** *Nat Protoc.* 2015; **10**(6): 845–858.
    PubMed Abstract | **Publisher Full Text**

13. Bordoli L, Kiefer F, Arnold K, *et al.*: **Protein structure homology modeling using SWISS-MODEL workspace.** *Nat Protoc.* 2009; **4**(1): 1–13.
    PubMed Abstract | **Publisher Full Text**

14. Costa MG, Benetti-Barbosa TG, Desdouits N, *et al.*: **Impact of M36I polymorphism on the interaction of HIV-1 protease with its substrates: insights from molecular dynamics.** *BMC Genomics.* 2014; **15**(Suppl 7): S5.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Kovalskyy D, Dubyna V, Mark AE, *et al.*: **A molecular dynamics study of the structural stability of HIV-1 protease under physiological conditions: the role of Na⁺ ions in stabilizing the active site.** *Proteins.* 2005; **58**(2): 450–8.
    PubMed Abstract | **Publisher Full Text**

16. Torbeev VY, Raghuraman H, Hamelberg D, *et al.*: **Protein conformational dynamics in the mechanism of HIV-1 protease catalysis.** *Proc Natl Acad Sci U S A.* 2011; **108**(52): 20982–7.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Referee Status: ☑ ☑ ❓

Version 2

Referee Report 24 June 2016

doi:10.5256/f1000research.9676.r14497

☑ **Saverio Alberti**
Unit of Cancer Pathology, Center of Excellence for Aging Sciences and Translational Medicine (CeSI-MeT), "G. d'Annunzio" University of Chieti-Pescara, Chieti, Italy

A clearer and more pointed presentation of key points is provided.

A broader discussion may begin on issues that remain open, such as large-scale crystallization efforts/ 'universal' protein modeling or on the role of mutations in non-coding DNA.

This may well form the basis for additional articles on this topic

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 23 June 2016

doi:10.5256/f1000research.9676.r14496

☑ **Attila Bérces**
Omixon, Inc., Budapest, Hungary

This revision significantly improved from the original article. Neoantigen prediction is a good example, where atomic level modelling can significantly contribute to clinical decisions.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 31 May 2016

Anita Grigoriadis

Breast Cancer Now Unit, Cancer Division, Faculty of Life Sciences & Medicine, King's College London, London, UK

In this article Oliver *et al.* highlight to importance of analytical methods such as "molecular modeling and molecular dynamic simulations" as tools that can significantly increase the interpretation of the functional impact of mutations/polymorphisms currently classified as variants of unknown significance (VUSs).

While the topic of molecular dynamics simulation for VUSs is worth discussing, the article could expand on

- the current limitation of these methods
- the complexity of these nucleotide changes in e.g. coding and non-coding genomic regions
- integration with other high throughput technologies and thus increase challenges for multimodal effects in such simulations
- HIV is given as an example, however other clinical examples, their associated specificity and limitations should be provided to make this article relevant for a wider audience.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Referee Report 18 May 2016

Attila Bérces

Omixon, Inc., Budapest, Hungary

The authors of this article "highlight a role for molecular modeling and molecular dynamic simulations as tools that can significantly increase the yield of information to aid in the evaluation of pathogenicity."

Molecular dynamics simulations are indeed excellent tools to study protein and nucleic acid structures, protein-ligand, protein-protein and protein-nucleic acid interactions. These structural changes and interactions are fundamental in molecular biology and thus molecular simulations are an excellent tool to increase our understanding of these biological processes. In particular the proposed method may help the understanding of coding variants associated with amino acid substitution. In addition, RNA structure can be modeled, which can be informative about some micro-RNA variants.

However, these methods are not without limitations and this paper does not state these limitations explicitly. Pathogenic effects come in all strength and complexities. Most often pathogenicity is not associated with a single variant, but with a combination of several variants and the effect of a particular variant depends on the genotypes of other genes. Most likely, variants will stay of unknown significance unless we develop a system level understanding of biology. We may reach a point that we can simulate viral functions at molecular level, but it will not be applied to human variants any time soon.

Molecular modelling has some limitations and requires expert knowledge to apply it properly.

While I agree that molecular modelling is an excellent tool to understand molecular biology, we need to develop a system level of understanding of the connection between molecular biology and genetics before we can apply molecular simulations. Currently, molecular modeling can make practical contribution to understand variants associated with strong effects or the molecular basis of some Mendelian disorders.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Referee Report 06 May 2016

doi:10.5256/f1000research.9254.r13600

Saverio Alberti

Unit of Cancer Pathology, Center of Excellence for Aging Sciences and Translational Medicine (CeSI-MeT), "G. d'Annunzio" University of Chieti-Pescara, Chieti, Italy

In this article the Authors seek ways on how to better interpret the functional impact of mutations/polymorphisms currently classified as variants of unknown significance (VUSs). To tackle such a problem, the authors propose "molecular modeling and molecular dynamic simulations as tools that can significantly increase the yield of information to aid in the evaluation of pathogenicity." Expectations are that this "will significantly benefit the mission of clinical genomics for individualized medicine."

The idea is good and deserves to be published. Most of all, it deserves to be extensively discussed.

Points to be discussed
As much as the idea is good, it would benefit from better highlighting its (current) limitations.

1. *Ab initio* molecular modeling is a difficult enterprise. It can rarely be performed to some extent of accuracy by automatized engines/softwares and often reaches good structure prediction in around 60% of cases.

   Several large-scale, tentatively whole-coding genome, protein crystallization initiatives have been undertaken. Discussion of the state of the art of these initiatives may provide metrics on what could be expected from real-life application of the Authors' proposal, and how this is expected to evolve in the near future.

2. Most mutations/polymorphisms are detected in non-coding regions of DNA. This is hardly surprising given that non-coding DNA represents approximately 97% of the total content of a mammalian cell. The authors mention mutations in non coding regions, but this deserves being more extensively discussed in the publication.

Suggestions

Discussion of the points above and of the suggestions below, may possibly lead to extend the proposed approach to additional layers of prediction.

Algorithms are there that do recognise promoter, enhancer and super-enhancer regions with increasing accuracy. Clearly, allocating a mutation/polymorphism in one such region would increase its investigational and possibly medical value.

Additional regions of interest are CpG islands upstream or, less reliably, within coding regions.

Additional regions of interest are anchoring sites of DNA loops to distinct nuclear/nuclear membrane regions.

Additional regions of interest are those hosting the increasing family of non-coding RNA (NCRNA), including long NCRNA and miRNA, that have been shown to play distinct regulatory functions and may have roles in cancer development.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Discuss this Article

Version 1

Author Response 13 Jun 2016

Gavin Oliver, Mayo Clinic in Rochester, USA

We wish to thank each of the reviewers for their thorough review of our article and their varied insightful comments and suggestions. We have considered all of these and both edited and expanded our argument to address them. Firstly, we better define the current norms within clinical genomics testing specifically, in order to better orient the reader and set the scenario for our opinions. Furthermore, the article now more clearly defines our argument in the context of the wider medical field, and better defines the scope of our suggested uses of protein modeling. We acknowledge the many, diverse approaches that likely represent future components of clinical testing and define our position in relation to these. Ultimately we feel that the comments received enable our new version to present our position with greater clarity and a more pointed argument, which is better tailored to a diversity of expert readers.

Competing Interests: No competing interests were disclosed.