Performance of mutation pathogenicity prediction tools on missense variants associated with 46,XY differences of sex development

Luciana R. Montenegro, Antônio M. Lerário, Miriam Y. Nishi, Alexander A.L. Jorge, Berenice B. Mendonca

OBJECTIVES: Single nucleotide variants (SNVs) are the most common type of genetic variation among humans. High-throughput sequencing methods have recently characterized millions of SNVs in several thousand individuals from various populations, most of which are benign polymorphisms. Identifying rare disease-causing SNVs remains challenging, and often requires functional in vitro studies. Prioritizing the most likely pathogenic SNVs is of utmost importance, and several computational methods have been developed for this purpose. However, these methods are based on different assumptions, and often produce discordant results. The aim of the present study was to evaluate the performance of 11 widely used pathogenicity prediction tools, which are freely available for identifying known pathogenic SNVs: Fathmm, Mutation Assessor, Protein Analysis Through Evolutionary Relationships (Phanter), Sorting Intolerant From Tolerant (SIFT), Mutation Taster, Polymorphism Phenotyping v2 (Polyphen-2), Align Grantham Variation Grantham Deviation (Align-GVGD), CAAD, Provean, SNPs&GO, and MutPred.

METHODS: We analyzed 40 functionally proven pathogenic SNVs in four different genes associated with differences in sex development (DSD): 17β-hydroxysteroid dehydrogenase 3 (HSD17B3), steroidogenic factor 1 (NR5A1), androgen receptor (AR), and luteinizing hormone/chorionic gonadotropin receptor (LHCGR). To evaluate the false discovery rate of each tool, we analyzed 36 frequent (MAF ≤ 0.01) benign SNVs found in the same four DSD genes. The quality of the predictions was analyzed using six parameters: accuracy, precision, negative predictive value (NPV), sensitivity, specificity, and Matthews correlation coefficient (MCC). Overall performance was assessed using a receiver operating characteristic (ROC) curve.

RESULTS: Our study found that none of the tools were 100% precise in identifying pathogenic SNVs. The highest specificity, precision, and accuracy were observed for Mutation Assessor, MutPred, SNP, and GO. They also presented the best statistical results based on the ROC curve statistical analysis. Of the 11 tools evaluated, 6 (Mutation Assessor, Phanter, SIFT, Mutation Taster, Polyphen-2, and CAAD) exhibited sensitivity > 0.90, but they exhibited lower specificity (0.42-0.67). Performance, based on MCC, ranged from poor (Fathmm=0.04) to reasonably good (MutPred=0.66).

CONCLUSION: Computational algorithms are important tools for SNV analysis, but their correlation with functional studies not consistent. In the present analysis, the best performing tools (based on accuracy, precision, and specificity) were Mutation Assessor, MutPred, and SNPs&GO, which presented the best concordance with functional studies.

KEYWORDS: Disorders of Sex Development; Pathogenicity Prediction Tools; Genetics.
to be associated with the DSD phenotype (3). Recent studies in individuals with DSDs have characterized numerous single nucleotide variants (SNV) in several genes, most of which are benign polymorphisms. However, distinguishing rare disease-causing SNVs from rare polymorphisms remains challenging. Functional studies for disease association variants are often used, but are laborious and time-consuming (4,5).

Many methods have been developed for the computational prediction of the pathogenicity of SNVs, which are based on evolutionary conservation, protein structure/function, or assembly parameters, such as allelic diversity, pathogenicity, and association with genome-wide association studies (6). Studies analyzing the performance of prediction programs have been completed using a large number of missense variants (7). In the present study, we compared the performance of 11 widely used pathogenic prediction tools in the analysis of proven pathogenic DSD-causing SNVs in four different genes.

### MATERIAL AND METHODS

#### Dataset

We analyzed 40 disease-causing SNVs in four different genes associated with DSD: 17ß-hydroxysteroid dehydrogenase 3 (HSD17B3), steroidogenic factor 1 (NR5A1), androgen receptor (AR), and luteinizing hormone/chorionic gonadotropin receptor (LHCGR). All pathogenic allele variants have been published with functional studies showing loss of function activity (Table 1). To evaluate the false discovery rate of each tool, we selected 36 frequent benign SNVs (MAF > 0.01) found in the same DSD genes (Table 1).

#### Prediction Methods

We selected 11 widely used pathogenic prediction tools freely available on the Web: Fathmm, Mutation Assessor, Protein Analysis Through Evolutionary Relationships (Phanter), SIFT (Sorting Intolerant From Tolerant), Mutation Taster, Polymorphism Phenotyping v2 (Polyphen-2), Align Grantham

---

Table 1 - Single Nucleotide Variants (SNV) in DSD-related genes used for prediction analysis.

| Gene/Protein                        | Pathogenic allelic Variant | Benign allelic variant | Reference |
|-------------------------------------|----------------------------|------------------------|-----------|
| HSD17B3                             | p.Ser65Leu                 | p.Val25Met             | (8)       |
| 17ß-hydroxysteroid dehydrogenase    | p.Arg80Gln                 | p.Val31Leu             | (8)       |
| ENST00000375263                     | p.Ala203Val                | p.Gly289Arg            | (9)       |
| NP_000188                           | p.Val205Glu                | p.Ile102Phe            | (8)       |
|                                     | p.Phe208Ile                | p.Glu114Lys            | (8)       |
|                                     | p.Glu215Asp                | p.Arg45Trp             | (8)       |
|                                     | p.Ser232Leu                | p.Arg45Gln             | (10)      |
|                                     | p.Met235Val                | p.Ser65Ala             | (10)      |
|                                     | p.Pro282Leu                | p.Ala223Val            | (8)       |
|                                     | p.Cys268Tyr                | p.Glu11Asp             | (11)      |
| NR5A1                               | p.Val15Met                 | p.Glu114Ala            | (12)      |
| Steroidogenic factor 1              | p.Val20Leu                 | p.Val173Met            | (12)      |
| ENST00000373588.8                   | p.His24Tyr                 | p.Val173Met            | (12)      |
| NP_004950                           | p.Arg39Pro                 | p.Gly178Arg            | (13)      |
|                                     | p.Met78Ile                 | p.Tyr211Cys            | (11)      |
|                                     | p.Gly91Ser                 | p.Pro235Leu            | (14)      |
|                                     | p.Pro235Leu                | p.Thr296Met            | (15)      |
|                                     | p.Trp279Arg                | p.Val355Met            | (13)      |
| AR                                  | p.Arg313Cys                | p.Leu437Gln            | (14)      |
| Androgen receptor                    | p.Cys579Phe                | p.Ala45Gly             | (16)      |
| ENST00000374690.8                   | p.Phe582Tyr                | p.Gln93Leu             | (16)      |
| NP_000035                           | p.Arg710Thr                | p.Gln7His              | (17)      |
|                                     | p.Gly724Asp                | p.Gln7His              | (18)      |
|                                     | p.Gly750Asp                | p.Gly216Arg            | (17)      |
|                                     | p.Ala765Thr                | p.Leu272Phe            | (17)      |
|                                     | p.Arg774His                | p.Leu341Met            | (17)      |
|                                     | p.Leu812Pro                | p.Val731Met            | (18)      |
|                                     | p.Arg855Cys                | p.Arg856Leu            | (17)      |
| LHCGR                               | p.Asp864Gly                | p.Ala37Thr             | (19)      |
| Luteinizing hormone/chorionic        | p.Ile374Thr*               | p.Ala45Gly             | (16)      |
| gonadotropin receptor               |                            |                        |           |
| ENST000000294954                    |                            |                        |           |
| NP_002224                           |                            |                        |           |

*Inactivating variants – phenotype: Leydig cell hypoplasia.

**Activating variants – phenotype: GIPP (gonadotropin-independent precocious puberty).
Table 2 - Basis of the in silico prediction algorithms.

| Program Name | URL and Key reference | Basis | Reference |
|--------------|-----------------------|-------|-----------|
| Fathmm       | http://fathmm.bioinformatics.org.uk/ | Evolutionary conservation | (28),(29),(30) |
| Mutation Assessor | http://mutationassessor.org/r3/ | Evolutionary conservation | (31) |
| Phanter      | http://www.pantherdb.org/ | Evolutionary conservation | (32) |
| SIFT         | https://sift.bioc.illinois.edu/ | Evolutionary conservation | (33) |
| Mutation Taster | http://www.mutationtaster.org/ | Protein structure/function and Evolutionary conservation | (34) |
| Polyphen-2   | http://genetics.bwh.harvard.edu/pph2/ | Protein structure/function and Evolutionary conservation | (35) |
| Align-GVGD   | http://agvgd.hci.utah.edu/ | Protein structure/function and Evolutionary conservation | (36) |
| MutPred      | http://mutpred.mutdb.org/index.html | Protein structure/function and Evolutionary conservation | (37) |
| CAAD         | https://cadd.gs.washington.edu/ | Protein structure/function and Evolutionary conservation | (38) |
| Provean      | http://provean.jcvi.org/index.php | Protein structure/function | (39) |
| SNPs&GO      | http://snps.biofold.org/snps-and-go/snps-and-go.html | Protein structure/function | (40),(41),(42) |

Table 3 - Predictions by algorithms.

| Type of SNV | Classification by the sites | Fathmm | Mutation Assessor | Phanter | SIFT | Mutation Taster | Polyphen-2 | Align-GVGD | MutPred | CAAD | Provean | SNPs&GO |
|------------|-----------------------------|--------|-------------------|--------|-----|----------------|-----------|-----------|---------|------|-------|--------|
| Pathogenic | Pathogenic                  | 35     | 38                | 38     | 36  | 37              | 37        | 33        | 33      | 36   | 32    | 32     |
| Benign     | Benign                      | 5      | 2                 | 2      | 4   | 3               | 3         | 7         | 7       | 4    | 8     | 8      |
| Pathogenic | Pathogenic                  | 27     | 14                | 21     | 16  | 17              | 13        | 28        | 6       | 12   | 11    | 6      |
| Benign     | Benign                      | 9      | 22                | 15     | 18  | 19              | 13        | 8         | 30      | 24   | 25    | 30     |

Table 4 - Performance of the prediction algorithms.

| Performance | Fathmm | Mutation Assessor | Phanter | SIFT | Mutation Taster | Polyphen-2 | Align-GVGD | MutPred | CAAD | Provean | SNPs&GO |
|-------------|--------|-------------------|--------|-----|----------------|-----------|-----------|---------|------|-------|--------|
| Accuracy    | 0.56   | 0.79              | 0.70   | 0.74| 0.74           | 0.76      | 0.54      | 0.83    | 0.79| 0.75   | 0.82   |
| Precision   | 0.56   | 0.73              | 0.64   | 0.69| 0.69           | 0.73      | 0.54      | 0.85    | 0.75| 0.74   | 0.84   |
| Specificity | 0.16   | 0.61              | 0.42   | 0.53| 0.53           | 0.53      | 0.22      | 0.83    | 0.67| 0.69   | 0.83   |
| Sensitivity | 0.88   | 0.95              | 0.95   | 0.92| 0.93           | 0.93      | 0.83      | 0.83    | 0.90| 0.80   | 0.80   |
| NPV         | 0.50   | 0.92              | 0.88   | 0.86| 0.86           | 0.84      | 0.53      | 0.81    | 0.86| 0.76   | 0.79   |
| MMC         | 0.04   | 0.60              | 0.44   | 0.50| 0.50           | 0.51      | 0.06      | 0.66    | 0.59| 0.50   | 0.63   |

Variation Grantham Deviation (Align-GVGD), CAAD, Provean, and SNPs&GO (Table 2).

Statistical Analysis

The quality of the predictions was analyzed using six parameters: accuracy, precision, negative predictive value (NPV), sensitivity, specificity, and Matthews correlation coefficient (MCC). In the equations below, tp, tn, fp, and fn refer to true positive, true negative, false positive, and false negative, respectively.

![Accuracy: Precision: NPV: Sensitivity: Specificity: MCC:](attachment:image)

The MCC (43) is an important statistics tool that is widely used in bioinformatics as a performance metric, as it is not affected by the differing proportions of neutral and pathogenic datasets predicted by the different programs. Additionally, we also assessed the overall performance of deleterious prediction with the receiver operating characteristic (ROC) curve and area under the curve (AUC), using MedCalc for Windows, version 15.0 (MedCalc Software, Ostend, Belgium). ROC curves are an indicator of probability and performance for classification problems at various threshold settings, and AUCs represent the degree or measure of separability. Together, they indicate how capable a model is of distinguishing between classes. The higher the AUC, the better the model is at predicting an outcome (44).

We randomized the results of prediction of SNV present in the DDS genes as pathogenic or benign, based on the classification given by each program. The classifications “probably benign”, “benign”, “low”, and “neutral”, provided by Fathmm, Mutation Assessor, Phanter, SIFT, Mutation Taster, Provean, and Polyphen-2, were considered as benign. The classifications “possibly damaging”, “probably damaging”, “high”, “medium”, “deleterious”, and “damaging”, were considered as pathogenic. For CAAD, which uses numeric scores, values ≤ 10 were classified as benign, and those > 10 were classified as pathogenic. Align-GVGD also used numeric scores, which were classified as benign up to 25, and anything above was classified as pathogenic.

RESULTS

Based on the results for each program, none of the tools were 100% precise in identifying pathogenic SNVs. The values for the parameters measured are listed in Table 3, and include all pathogenic and benign variants. Phanter had the highest precision in the classification of pathogenic variants (38 out of 40 known to be pathogenic), followed by Mutation Taster and Polyphen-2 (both 37 out of 40 known to be pathogenic). Align-GVGD correctly classified fewer known pathogenic SNVs than any other tool (33 of 40 known to be pathogenic). Phanter and Mutation Taster both classified a high number of know benign SNVs as pathogenic (21 and 17, respectively, of 36).

Mutation Assessor, MutPred, and SNPs&GO presented more consistent results regarding the nature of the SNVs
DISCUSSION

The comparative predictive performance of each tool was evaluated using the AUC scores from ROC plots and the true negative rate (TNR, or specificity) as measurements. The analysis was separated into random groups, since the program analyzed a maximum of six samples at a time (Figure 1). The AUC values varied from low (Mutation Taster=0.55) to reasonably good (SNPs-&-GO=0.89), and two programs (Fathmn and Align-GVGD) for which the MMC values were reasonably good (SNPs-&-GO=0.89), and two programs (Fathmn and Align-GVGD) for which the MMC values were poor (0.04 and 0.06, respectively), improved in the statistical analysis made using the ROC curve (0.67 and 0.57, respectively). Fathmn and Align-GVGD exhibited the worst performance, with a high number of false positive results (MMC=0.04 and 0.06, respectively).

The highest-performing tools were Mutation Assessor, MutPred, and SNPs&GO.

ACKNOWLEDGMENTS

This work was supported by Grant 2013/02162-8 (to B.B.M) from the Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) and Grant 455845/2018-4 (to B.B.M) from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

AUTHOR CONTRIBUTIONS

Montenegro LR contributed to the acquisition, analysis, interpretation of data, and drafting of the article. Lerario AM contributed to the interpretation of data and reviewing the article. Nishi MY contributed to the interpretation of data, drafting and revising the article. Jorge AA contributed to the analysis and interpretation of data. Mendonca BB contributed to the conception and design of the study, drafting and revising the article.

REFERENCES

1. Wisierniewski AB, Batista RL, Costa EMF, Finlayson C, Sircili MHP, Dênes FT, et al. Management of 46,XY Differences/Disorders of Sex Development (DSD) Throughout Life. Endocr Rev. 2019;40(6):1547-72. https://doi.org/10.1210/er.2019-00049
2. Lee PA, Houk CP, Ahmed SF, Hughes IA; International Consensus Conference on Intersex organized by the Lawson Wilkins Pediatric Endocrine Society and the European Society for Paediatric Endocrinology. Consensus statement on management of intersex disorders. International Consensus Conference on Intersex. Pediatrics. 2006;118(2):e488-500. https://doi.org/10.1542/peds.2006-0738
3. Mendonca BB. Gender assignment in patients with disorder of sex development. Curr Opin Endocrinol Diabetes Obes. 2014;21(6):531-4. https://doi.org/10.1097/MED.0000000000000115
4. Kolesinska Z, Acierno J Jr, Ahmed SF, Xu C, Kapczuk K, Skorczyk-Werner A, et al. Integrating clinical and genetic approaches in the diagnosis of 46, XY disorders of sex development. Endocr Connect. 2018;7(12):1480-90. https://doi.org/10.1530/EC-18-0472
5. Hughes LA, McKay-Bounford K, Webb EA, Dasani P, Clokie S, Chandran H, et al. Next generation sequencing (NGS) to improve the diagnosis and management of patients with disorders of sex development (DSD). Endocr Connect. 2019;8(2):100-10. https://doi.org/10.1530/EC-18-0376
6. Leong IU, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. BMC Med Genet. 2015;16:34. https://doi.org/10.1186/s12881-015-0176-z
7. Thuesberg J, Olaltubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat. 2011;32(4):358-68. https://doi.org/10.1002/humu.21445
8. Andersson S, Geissler WM, Wu L, Davis DL, Grumbach MM, New MI, et al. Molecular genetics and pathophysiology of 17 beta-hydroxysteroid dehydrogenase 3 deficiency. J Clin Endocrinol Metab. 1996;81(1):130-6. https://doi.org/10.1210/jcem.81.1.8550739
9. Dong C, Wei P, Jian X, Gibbs R, Boerwinkel E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for non-synonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015;24(8):2125-37. https://doi.org/10.1093/hmg/ddu233

10. Ceolser WM, Davis DL, Wu L, Bradshaw KD, Patel S, Mendonca BB, et al. Male pseudohermaphroditism caused by mutations of testicular 17β-hydroxysteroid dehydrogenase 3. Nat Genet. 1994;7(1):34-9. https://doi.org/10.1038/ng0594-34

11. Lindqvist A, Hughes IA. Anderson S. Substitution mutation C268Y causes 17β-hydroxysteroid dehydrogenase 3 deficiency. J Clin Endocrinol Metab. 2001;86(2):921-3. https://doi.org/10.1210/jcem.86.2.7172

12. Lin L, Philbert P, Ferraz-de-Souza B, Kelberman D, Homfray T, Albanese A, et al. Heterozygous missense mutations in steroidogenic factor 1 (SF1/Ad4BP/NR5A1) are associated with 46,XY disorders of sex development with normal adrenal function. J Clin Endocrinol Metab. 2007;92(3):991-9. https://doi.org/10.1210/jc.2006-1672

13. Camats N, Pandey AV, Fernández-Cancio M, Andaluz P, Janner M, Torán N, et al. Ten novel mutations in the NR5A1 gene caused disorder sex development in 46,XY and ovarian insufficiency in 46,XX individuals. J Clin Endocrinol Metab. 2012;97(7):E1294-306. https://doi.org/10.1210/jc.2012-13169

14. Allali S, Muller JB, Brauner R, Lourenc o D, Boudjenah R, Karageorgou V, et al. Mutation analysis of NR5A1 encoding steroidogenic factor 1 in 77 patients with 46, XY disorders of sex development (DSD) including hypoplasia. PLoS One. 2011;6(10):e24117. https://doi.org/10.1371/journal.pone.0024117

15. Lin L, Achermann JC. Steroidogenic factor-1 (SF-1, Ad4BP/NR5A1) and disorders of tests development. Sex Dev. 2008;8(4-5):200-9. https://doi.org/10.1159/000152536

16. Imasaki K, Okabe T, Murakami H, Tanaka Y, Haji M, Takayanagi R, et al. Androgen insensitivity syndrome due to new mutations in the DNA-binding domain of the androgen receptor. Mol Cell Endocrinol. 1996;120(1):15-24. https://doi.org/10.1016/S0303-7207(96)03812-9

17. Ahmed SF, Cheng A, Dovely L, Hawkins JR, Martin H, Rowland J, et al. Phenotypic features, androgen receptor binding, and mutational analysis in 278 clinical cases reported as androgen insensitivity syndrome. J Clin Endocrinol Metab. 2005;88(2):638-65. https://doi.org/10.1210/jcem.2005.6.26337

18. Jääskeläinen J, Mongan NP, Harland S, Hughes IA. Five novel androgen receptor gene mutations associated with complete androgen insensitivity syndrome. Hum Mutat. 2006;27(3):291. https://doi.org/10.1002/humu.9405

19. Palo-Ryöaasräm R, Liu G, Brickman W, Duranteau L, Monroe J, El-Awdy MK, et al. A novel double mutation in the luteinizing hormone receptor in a kindred with familial Leydig cell hypoplasia and male pseudohermaphroditism. Endocrinol Res. 2005;31(4):307-23. https://doi.org/10.1080/0365965050430890

20. Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. Endocr Rev. 2000;21(5):551-83. https://doi.org/10.1210/edrv.21.5.0409

21. Stavrov SS, Zhu YS, Cai LJ, Katz MD, Herrera C, Defilippo-Ricart M, et al. A novel mutation of the human luteinizing hormone receptor in 46XY and 46XX sisters. J Clin Endocrinol Metab. 1998;83(6):2091-8. https://doi.org/10.1210/jcem.83.6.8585

22. Leung MY, Ananth H, Wu O, Wu SM, Aziz A, Inam S, Awadh M, et al. A novel missense homozygous inactivating mutation in the fourth transmembrane helix of the luteinizing hormone receptor in Leydig cell hypoplasia. Am J Med Genet A. 2004;130A(2):146-53. https://doi.org/10.1002/ajmg.a.20681

23. Krawczak M, Birnstiel ML, Montenegro LR et al. Inferring the molecular and phenotypic impact of amino acid variants using position-specific evolutionary preservation. Bioinformatics. 2012;28(14):2230-2. https://doi.org/10.1093/bioinformatics/bts222

24. Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. Genome Biol. 2007;8(11):R232. https://doi.org/10.1186/gb-2007-8-11-r232

25. Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet. 2006;43(4):295-305. https://doi.org/10.1136/jmg.2005.033878

26. Laue L, Wu SM, Kudo M, Husea AJ, Cutler GB Jr, Jolly DH, et al. Heterogeneity of activating mutations of the human luteinizing hormone receptor gene in familial male-limited precocious puberty. Proc Natl Acad Sci U S A. 1995;92(6):1906-10. https://doi.org/10.1073/pnas.92.6.1906

27. Latronico AC, Anasti J, Arnhold IJ, Mendonça BB, Domenice S, Albano MC, et al. A novel mutation of the luteinizing hormone receptor gene causing male gonadotropin-independent precocious puberty. J Clin Endocrinol Metab. 1995;80(8):2490-4. https://doi.org/10.1210/jeem.80.8.7629248

28. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013;34(1):57-65. https://doi.org/10.1002/humu.22225

29. Stavrov SS, Zhu YS, Cai LJ, Katz MD, Herrera C, Defilippo-Ricart M, et al. A unique constitutively activating mutation in third transmembrane helix of luteinizing hormone receptor causes sporadic male gonadotropin-independence precocious puberty. J Clin Endocrinol Metab. 1998;83(7):2435-40. https://doi.org/10.1210/jcem.83.7.4986

30. Laue L, Chan WY, Husea AJ, Kudo M, Husea SY, Wu SM, et al. Genetic heterogeneity of constitutively activating mutations of the human