

## COMMENT

## One is the loneliest number: genotypic matchmaking using the electronic health record

Elly Brokamp<sup>1</sup>✉, Mary E. Koziura<sup>1</sup>, John A. Phillips III<sup>1</sup>, Leigh Anne Tang<sup>3</sup>, Joy D. Cogan<sup>1</sup>, Lynette C. Rives<sup>1</sup>, Amy K. Robertson<sup>1</sup>, Laura Duncan<sup>1</sup>, Anna Bican<sup>1</sup>, Josh F. Peterson<sup>3</sup>, John H. Newman<sup>2</sup>, Rizwan Hamid<sup>1</sup>, Lisa Bastarache<sup>3</sup>, Undiagnosed Diseases Network\*

*Genetics in Medicine* (2021) 23:1830–1832; <https://doi.org/10.1038/s41436-021-01179-w>

When genetic testing detects a candidate variant (CV) in a gene that is not known to cause disease in humans, clinicians face a dilemma. Where can they find the additional evidence needed to make a diagnosis? One action a clinician can take is to search for additional patients with a deleterious variant in the same gene. Here we present a case study where the electronic health record (EHR) was used to facilitate genotypic matchmaking to help diagnose a patient at Vanderbilt's Undiagnosed Disease Network (UDN)<sup>1</sup> clinical site.

A 26-year-old female with unexplained mild intellectual disability (ID), autism spectrum disorder (ASD), obsessive compulsive disorder (OCD), high myopia, and joint hypermobility had exome sequencing (ES) as part of her work-up for the UDN. Interestingly, a de novo *MSL2* p.S232Tfs\*10 variant was included in the ES report along with its possible association with ASD.<sup>2</sup>

The *MSL2* protein had been reported to be part of a complex that plays a major role in chromatin regulation and structure through histone H4 acetylation and H2B ubiquitylation.<sup>3,4</sup> Due to the participant's CV being a frameshift, de novo, and the function of the *MSL2* protein, the Vanderbilt University Medical Center (VUMC) UDN team considered it an attractive candidate, but the lack of a previously reported role in Mendelian disease precluded it from being diagnostic.

The UDN's process to find additional cases is to upload CVs to PhenomeCentral, a founding member of Matchmaker Exchange.<sup>5</sup> A widely used, similar strategy is for the provider to post the candidate gene in GeneMatcher, MyGene2, and/or Matchmaker Exchange.<sup>6–8</sup> These online connection tools rely on other providers using these platforms, require some length of waiting time, and need communication between providers to share additional information to possibly solidify a match. To date, the VUMC team has not had any *MSL2* matches through PhenomeCentral.

Having failed to find a match through available matchmaking platforms, the Vanderbilt UDN team searched for "*MSL2*" in VUMC's Synthetic Derivative (SD), a continuously updated, de-identified database of EHRs for more than 3 million patients.<sup>9</sup> This revealed two patients whose charts were subsequently reviewed using SD Discover, a browser-based application developed at Vanderbilt to facilitate chart review. Both patients were found to have de novo *MSL2* variants identified through clinical trio ES and had strikingly similar phenotypes to the proband.

The first patient identified through SD was a 15-year-old male with a history of ASD, mild ID, obsessive-compulsive disorder (OCD), hypermobile Ehlers-Danlos syndrome (EDS), and high myopia. He was heterozygous for a de novo p.Pro25Ser *MSL2* CV, a

well conserved variant that is predicted to be damaging based on in silico models. The second patient was a 13-year-old female with a history of global developmental delay, ASD, attention deficit disorder (ADD), visual and language processing disorder, OCD, high myopia, and hypermobile EDS. She was heterozygous for a de novo p.Thr217AspfsX2 *MSL2* CV, a variant that created a premature stop codon and was predicted to cause loss of normal protein function. The variants found in both patients were absent from large population cohorts.

In both cases, clinical genetics providers were suspicious of the *MSL2* CV as a cause of their patients' ASD, behavioral problems, and possibly their hypermobility. They documented that there was not enough evidence to consider the *MSL2* CV to be diagnostic.

All three individuals had a de novo, likely gene disrupting, *MSL2* CV. Since all three individuals also had similar phenotypes (ASD, OCD, mild ID or learning disabilities, and an EDS type III or joint hypermobility), this suggests its causative role in a new disorder. These additional patients gave the team the confidence to initiate contact with clinical laboratories to find more individuals with *MSL2* as a candidate gene and connect with a team doing functional studies. Finding these two additional individuals was an integral part of giving the UDN participant a molecular diagnosis for a new disease.

The American College of Medical Genetics and Genomics (ACMG) recently published a statement encouraging providers to optimize their use of genomic testing results in the EHR.<sup>10</sup> Based on the experience with *MSL2*, VUMC's team proposes another use for genomic testing results in the EHR: a full text search for a gene name to quickly ascertain additional individuals with shared phenotype(s) and candidate gene(s). We have termed this strategy "EHR-based genotype matchmaking" to indicate the process of finding a genotype match that has been clinically tested and might provide evidence for causality.

EHR-based genotype matchmaking offers several benefits beyond what is realized by current methods. EHRs contain dense and longitudinal medical histories that are not always available to clinical laboratories; also, EHRs contain a large and growing population of patients who have received genetic testing compared with patient cohorts whose data have been uploaded to matchmaking platforms. Motivated by the successful resolution of the case above, the Vanderbilt UDN team has added EHR-based genotype matchmaking to its variant analysis pipeline and is in the process of reanalyzing past cases.

The VUMC UDN team recognizes that this powerful tool has potential barriers that complicate implementation. Like many medical centers, VUMC does not currently store genetic data in an

<sup>1</sup>Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>2</sup>Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>3</sup>Center for Precision Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. \*A list of authors and their affiliations appears at the end of the paper. ✉email: elly.brocamp@vumc.org

Received: 7 August 2020; Accepted: 5 April 2021;  
Published online: 6 July 2021

## REFERENCES

- Gahl, W., Wise, A. & Ashley, E. The Undiagnosed Diseases Network of the National Institutes of Health: a national extension. *JAMA*. **314**, 1797–1798 (2015).
- Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. **515**, 216–221 (2014).
- Smith, E. R. et al. A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol. Cell. Biol.* **25**, 9175–9188 (2006).
- Wu, L., Zee, B. M., Wang, Y., Garcia, B. A. & Dou, Y. The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. *Mol. Cell*. **43**, 132–144 (2011).
- Buske, O. K. et al. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum. Mutat.* **36**, 931–940 (2015).
- Philippakis, A. A. et al. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum. Mutat.* **36**, 915–921 (2015).
- Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).
- NHGRI/NHLBI University of Washington–Center for Mendelian Genomics (UW-CMG). MyGene2. <http://www.mygene2.org/> (2020).
- Roden, D. M. et al. Development of a large-scale de-identified DNA biobank to Enable Personalized Medicine. *Clin Pharmacol Ther* **84**, 362–369 (2008).
- Grebe, T. et al. The interface of genomic information with electronic health record: a points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **22**, 1431–1436. <https://doi.org/10.1038/s41436-020-0841-2> (2020).
- Appelbaum, P. S. et al. Is there a duty to reinterpret genetic data? The ethical dimensions. *Genet. Med.* **22**, 633–639 (2020).
- David, K. L. et al. Patient re-contact after revision of genomic test results: points to consider—a statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **21**, 769–771 (2019).
- Farwell-Hagman, K. et al. Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. *Genet. Med.* **19**, 224–235 (2017).
- Retterer, M. et al. Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* **18**, 696–704 (2015).

## ACKNOWLEDGEMENTS

This research was supported by the National Institutes of Health (NIH) Common Fund, through the Office of Strategic Coordination/Office of the NIH Director under award number U01HG007672. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to E.B.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

easily accessible format. Genetic testing reports are mainly uploaded to the EHR as PDF files that are currently not searchable in the text-based SD. Our geneticists and genetic counselors record genetic results as text in the EHR, but this is a time-consuming process that is subject to typos and errors, and not all VUMC providers record genetic test results in this way. Fortunately, many genes have fairly distinctive names—the search for “MSL2” easily found the two records above. However, searching for gene names that are short (e.g., *TR*), overlap with common terms, or that are frequently reported as somatic variants (e.g., *EGFR*) can yield thousands of irrelevant records.

Major EHR vendors are beginning to develop and market ancillary systems and modules to store structured genetic test results. If adopted, these new features, along with standard reporting of results from laboratories, will greatly simplify the process of identifying patients with matching genotypes. In the meantime, the UDN team plans to continue to index medical records based on a string-matching approach for new cases, and to reanalyze previous undiagnosed cases. The database that stores these string matches has become an invaluable resource for rapidly assessing new UDN cases, and the growth illustrates the rapid growth over the last decade of genetic data in the EHR.

More work is needed to streamline the dissemination of knowledge gained from EHR-based patient matching. The Vanderbilt UDN team is working on an institutional review board (IRB) application that would allow them to inform the ordering provider of the new connections regarding their patients' genetic testing results. While the research team is not the clinical laboratory who performed the testing, there is still an ethical duty to reinterpret unsolved exomes and inform the ordering provider/patients of new conclusions.<sup>11,12</sup> This IRB permission would allow the research team to also ask clarifying phenotype questions that are unanswered in the EHR. Ideally, VUMC's search process, which is done on a research basis, could potentially be done as part of routine clinical care in the future.

Further work may be directed toward more automated phenotyping methods that use the EHR. The combination of structured genetic data and automated phenotyping methods may facilitate more automated approaches to scaling EHR-based matchmaking, similar to approaches used by matchmaking platforms.

The problem with candidate genes is not restricted to a research setting. Clinical ES, which is increasing in popularity, often reports CVs to clinicians. Although it may not be standard for all clinical laboratories to report variants in candidate genes, two large clinical laboratories have reported that ~8% of their diagnostic ES testing reports a candidate gene.<sup>13,14</sup> One laboratory shared that a candidate gene was included on a total of 24% of their reports.<sup>14</sup> Linking candidate genes found in these tests with the medical phenome described in the EHR represents an opportunity to improve interpretation of exomes, which will in turn improve patient care, increase the utility of clinical genetic testing, and further the discovery of new Mendelian diseases. The institutional and technological changes necessary to enable this function in other institutions will take extensive effort and funding, but we hope that this report creates additional incentive for change.

## UNDIAGNOSED DISEASES NETWORK

Mercedes E. Alejandro<sup>4</sup>, Mahshid S. Azamian<sup>4</sup>, Carlos A. Bacino<sup>4</sup>, Ashok Balasubramanyam<sup>4</sup>, Bret L. Bostwick<sup>4</sup>, Lindsay C. Burrage<sup>4</sup>, Hsiao-Tuan Chao<sup>4</sup>, Gary D. Clark<sup>4</sup>, William J. Craigen<sup>4</sup>, Shweta U. Dhar<sup>4</sup>, Lisa T. Emrick<sup>4</sup>, Alica M. Goldman<sup>4</sup>, Neil A. Hanchard<sup>4</sup>, Fariha Jamal<sup>4</sup>, Lefkothea Karaviti<sup>4</sup>, Seema R. Lalani<sup>4</sup>, Brendan H. Lee<sup>4</sup>, Richard A. Lewis<sup>4</sup>, Ronit Marom<sup>4</sup>, Paolo M. Moretti<sup>4</sup>, David R. Murdock<sup>4</sup>, Sarah K. Nicholas<sup>4</sup>, James P. Orenge<sup>4</sup>, Jennifer E. Posey<sup>4</sup>, Lorraine Potocki<sup>4</sup>, Jill A. Rosenfeld<sup>4</sup>, Susan L. Samson<sup>4</sup>, Daryl A. Scott<sup>4</sup>, Alyssa A. Tran<sup>4</sup>, Tiphonie P. Vogel<sup>4</sup>, Hugo J. Bellen<sup>5</sup>, Michael F. Wangler<sup>5</sup>, Shinya Yamamoto<sup>5</sup>, Christine M. Eng<sup>6</sup>, Pengfei Liu<sup>6</sup>, Patricia A. Ward<sup>6</sup>, David B. Goldstein<sup>7</sup>, Nicholas Stong<sup>7</sup>, Heidi Cope<sup>8</sup>, Yong-hui Jiang<sup>8</sup>, Allyn McConkie-Rosell<sup>8</sup>,

Kelly Schoch<sup>8</sup>, Vandana Shashi<sup>8</sup>, Rebecca C. Spillmann<sup>8</sup>, Jennifer A. Sullivan<sup>8</sup>, Queenie K.-G. Tan<sup>8</sup>, Nicole M. Walley<sup>8</sup>, Pankaj Agrawal<sup>9</sup>, Alan H. Beggs<sup>9</sup>, Gerard T. Berry<sup>9</sup>, Lauren C. Briere<sup>9</sup>, Cynthia M. Cooper<sup>9</sup>, Laurel A. Cobban<sup>9</sup>, Elizabeth L. Fieg<sup>9</sup>, Frances High<sup>9</sup>, Ingrid A. Holm<sup>9</sup>, Susan Korrick<sup>9</sup>, Joel B. Krier<sup>9</sup>, Sharyn A. Lincoln<sup>9</sup>, Joseph Loscalzo<sup>9</sup>, Richard L. Maas<sup>9</sup>, Calum A. MacRae<sup>9</sup>, J. Carl Pallais<sup>9</sup>, Lance H. Rodan<sup>9</sup>, Edwin K. Silverman<sup>9</sup>, Joan M. Stoler<sup>9</sup>, David A. Sweetser<sup>9</sup>, Melissa Walker<sup>9</sup>, Chris A. Walsh<sup>9</sup>, Cecilia Esteves<sup>10</sup>, Emily G. Kelley<sup>10</sup>, Isaac S. Kohane<sup>10</sup>, Kimberly LeBlanc<sup>10</sup>, Alexa T. McCray<sup>10</sup>, Matthew Might<sup>10</sup>, David P. Bick<sup>11</sup>, Camille L. Birch<sup>11</sup>, Braden E. Boone<sup>11</sup>, Donna M. Brown<sup>11</sup>, Daniel C. Dorset<sup>11</sup>, Angela L. Jones<sup>11</sup>, Jozef Lazar<sup>11</sup>, Shawn E. Levy<sup>11</sup>, Thomas May<sup>11</sup>, J. Scott Newberry<sup>11</sup>, Elizabeth A. Worthey<sup>11</sup>, Ian R. Lanza<sup>12</sup>, Devin Oglesbee<sup>12</sup>, Eva Morava-Kozicz<sup>12</sup>, Brendan C. Lanpher<sup>12</sup>, Surendra Dasari<sup>12</sup>, Guney Bademci<sup>13</sup>, Deborah Barbouth<sup>13</sup>, Stephanie Bivona<sup>13</sup>, Olveen Carrasquillo<sup>13</sup>, Ta Chen Peter Chang<sup>13</sup>, Irman Forghani<sup>13</sup>, Alana Grajewski<sup>13</sup>, Rosario Isasi<sup>13</sup>, Byron Lam<sup>13</sup>, Roy Levitt<sup>13</sup>, Xue Zhong Liu<sup>13</sup>, Jacob McCauley<sup>13</sup>, Ralph Sacco<sup>13</sup>, Mario Saporta<sup>13</sup>, Judy Schaechter<sup>13</sup>, Mustafa Tekin<sup>13</sup>, Fred Telischi<sup>13</sup>, Willa Thorson<sup>13</sup>, Stephan Zuchner<sup>13</sup>, Gabriel F. Batzli<sup>14</sup>, Heather A. Colley<sup>14</sup>, Jyoti G. Dayal<sup>14</sup>, David J. Eckstein<sup>14</sup>, Donna M. Krasnewich<sup>14</sup>, Laura A. Mamounas<sup>14</sup>, Teri A. Manolio<sup>14</sup>, John J. Mulvihill<sup>14</sup>, Robb K. Rowley<sup>14</sup>, Catherine H. Sillari<sup>14</sup>, Cecelia P. Tamburro<sup>14</sup>, Tiina K. Urv<sup>14</sup>, Anastasia L. Wise<sup>14</sup>, Maria T. Acosta<sup>15</sup>, Carsten Bonnenmann<sup>15</sup>, Elizabeth A. Burke<sup>15</sup>, Precilla D'Souza<sup>15</sup>, Mariska Davids<sup>15</sup>, David D. Draper<sup>15</sup>, Tyra Estwick<sup>15</sup>, Carlos Ferreira<sup>15</sup>, Rena A. Godfrey<sup>15</sup>, Catherine A. Groden<sup>15</sup>, Jean M. Johnston<sup>15</sup>, C. Christopher Lau<sup>15</sup>, Ellen F. Macnamara<sup>15</sup>, Valerie V. Maduro<sup>15</sup>, Thomas C. Markello<sup>15</sup>, Marie Morimoto<sup>15</sup>, Avi Nath<sup>15</sup>, Donna Novacic<sup>15</sup>, Barbara N. Pusey<sup>15</sup>, Prashant Sharma<sup>15</sup>, Camilo Toro<sup>15</sup>, Colleen E. Wahl<sup>15</sup>, Guoyun Yu<sup>15</sup>, Andrea L. Gropman<sup>16</sup>, Eva Baker<sup>17</sup>, David R. Adams<sup>18</sup>, William A. Gahl<sup>18</sup>, May Christine V. Malicdan<sup>18</sup>, Cynthia J. Tiffit<sup>18</sup>, Lynne A. Wolfe<sup>18</sup>, John Yang<sup>18</sup>, Matthew Brush<sup>19</sup>, Jean-Philippe F. Gourdine<sup>19</sup>, Melissa Haendel<sup>19</sup>, David M. Koeller<sup>19</sup>, Jennifer E. Kyle<sup>20</sup>, Thomas O. Metz<sup>20</sup>, Katrina M. Waters<sup>20</sup>, Bobbie-Jo M. Webb-Robertson<sup>20</sup>, Euan A. Ashley<sup>21</sup>, Jonathan A. Bernstein<sup>21</sup>, Devon Bonner<sup>21</sup>, Terra R. Coakley<sup>21</sup>, Liliana Fernandez<sup>21</sup>, Paul G. Fisher<sup>21</sup>, Jason Hom<sup>21</sup>, Yong Huang<sup>21</sup>, Jennefer N. Kohler<sup>21</sup>, Marta M. Majcherska<sup>21</sup>, Shruti Marwaha<sup>21</sup>, Colleen E. McCormack<sup>21</sup>, Chloe M. Reuter<sup>21</sup>, Jacinda B. Sampson<sup>21</sup>, Kevin S. Smith<sup>21</sup>, Matthew T. Wheeler<sup>21</sup>, Diane B. Zastrow<sup>21</sup>, Chunli Zhao<sup>21</sup>, Maura Ruzhnikov<sup>21</sup>, Laure Fresard<sup>21</sup>, Holly K. Tabor<sup>21</sup>, Archana N. Raja<sup>21</sup>, Gill Bejerano<sup>21</sup>, Beth A. Martin<sup>21</sup>, Shirley Sutton<sup>21</sup>, Patrick Allard<sup>22</sup>, Manish J. Butte<sup>22</sup>, Esteban C. Dell'Angelica<sup>22</sup>, Naghmeh Dorrani<sup>22</sup>, Emilie D. Douine<sup>22</sup>, Brent L. Fogel<sup>22</sup>, Alden Huang<sup>22</sup>, Deborah Krakow<sup>22</sup>, Hane Lee<sup>22</sup>, Sandra K. Loo<sup>22</sup>, Martin G. Martin<sup>22</sup>, Julian A. Martinez-Agosto<sup>22</sup>, Stanley F. Nelson<sup>22</sup>, Christina G. S. Palmer<sup>22</sup>, Jeanette C. Papp<sup>22</sup>, Neil H. Parker<sup>22</sup>, Genecee Renteria<sup>22</sup>, Rebecca H. Signer<sup>22</sup>, Janet S. Sinsheimer<sup>22</sup>, Jijun Wan<sup>22</sup>, Lee-kai Wang<sup>22</sup>, Jeremy D. Woods<sup>22</sup>, Amanda J. Yoon<sup>22</sup>, Lorenzo Botto<sup>23</sup>, Ashley Andrews<sup>23</sup>, Nicola Longo<sup>23</sup>, John Carey<sup>23</sup>, John Bohnsack<sup>23</sup>, Dave Viskochil<sup>23</sup>, Moretti Paolo<sup>23</sup>, Pace Laura<sup>23</sup>, Justin Alvey<sup>23</sup>, Jim Bale<sup>23</sup>, Gabor Marth<sup>23</sup>, Matt Velinder<sup>23</sup>, Aaron Quinlan<sup>23</sup>, Rong Mao<sup>24</sup>, Pinar Bayrak-Toydemir<sup>24</sup>, John H. Postlethwait<sup>25</sup>, Monte Westerfield<sup>25</sup>, Lisa Shakachite<sup>26</sup>, F. Sessions Cole<sup>27</sup>, Nichole Hayes<sup>27</sup>, Dana Kiley<sup>27</sup>, Kathleen Shields<sup>27</sup>, Kathy Sisco<sup>27</sup>, Jennifer Wambach<sup>27</sup>, Daniel Wegner<sup>27</sup>, Dustin Baldrige<sup>27</sup>, Timothy Schedl<sup>27</sup>, Lilianna Solnica-Krezel<sup>27</sup>, Stephen Pak<sup>27</sup> and Jimann Shin<sup>27</sup>

<sup>4</sup>BCM Clinical, Houston, TX, USA. <sup>5</sup>BCM MOSC, Houston, TX, USA. <sup>6</sup>BCM Sequencing, Houston, TX, USA. <sup>7</sup>Columbia, New York City, NY, USA. <sup>8</sup>Duke Nashville, Durham, NC, USA. <sup>9</sup>Harvard, Cambridge, MA, USA. <sup>10</sup>Harvard CC, Cambridge, MA, USA. <sup>11</sup>HudsonAlpha, Huntsville, AL, USA. <sup>12</sup>Mayo Clinic, Rochester, MN, USA. <sup>13</sup>University of Miami, Miami, FL, USA. <sup>14</sup>NIH, Bethesda, MD, USA. <sup>15</sup>NIH UDP, Bethesda, MD, USA. <sup>16</sup>NIH UDP, CNMC, Bethesda, MD, USA. <sup>17</sup>NIH UDP, DRM, Bethesda, MD, USA. <sup>18</sup>NIH UDP, NHGRI, Bethesda, MD, USA. <sup>19</sup>OHSU Metabolomics, Portland, OR, USA. <sup>20</sup>PNNL Metabolomics, Richland, WA, USA. <sup>21</sup>Stanford University, Stanford, CA, USA. <sup>22</sup>UCLA, Los Angeles, CA, USA. <sup>23</sup>University of Utah, Salt Lake City, UT, USA. <sup>24</sup>University of Utah/ARUP, Salt Lake City, UT, USA. <sup>25</sup>UO MOSC, Eugene, OR, USA. <sup>26</sup>Vanderbilt University Medical Center, Nashville, TN, USA. <sup>27</sup>Washington University Clinical, St. Louis, MO, USA.