



Second European CMT Specialists Conference



PUBLICATIONS/READINGS:

Health data management and the use of AI in (iNMD) diagnostics (Wolfgang M. Pernice) M 10 / WP 3

Readings for preparing:

Plenary Session 2: Methods: diagnostics/genetics of CMT neuropathies
Clinical trials, data sharing and outcome measurement

I. AI in iNMD Diagnostics:

1. **Variant Prediction Models:** A major topic in iNMDs remains the (genetic) diagnostic gap, i.e. the fact that in about ca. 50% of e.g. CMT cases, we are unable to determine the genetic cause of the disease. Typically, this is because we are unable to predict whether or not a particular variant is pathogenic or not. Such variants of uncertain significance (VUS) may fall into known CMT genes (not all variants in a given CMT gene are damaging) or may fall into novel genes not (yet) associated with CMT. AI

systems play an increasingly important role in trying to predict whether or not a given VUS is pathogenic. Such models are typically called Variant Effect Prediction (VEP) models. Several recent advances are noticeable. While many models produce seemingly impressive results on benchmarks, their utility in practice remains largely unproven.

2. **AlphaMissense** (Google DeepMind, 2023): Leveraging AlphaFold2, a leading protein-structure prediction model (Nobel-prize in 2025), to predict the impact of missense variants.
 - **Full citation:** Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T. and Schneider, R.G., 2023. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, 381(6664), p.eadg7492.
Link: <https://www.science.org/doi/10.1126/science.adg7492>
3. **EVE (Cheng et. al 2023)**: A generative model of variant effect that does not use disease labels (i.e. information on whether a particular variant in the training set is pathogenic or not).
 - **Full citation:** Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J.K., Brock, K., Gal, Y. and Marks, D.S., 2021. Disease variant prediction with deep generative models of evolutionary data. *Nature*, 599(7883), pp.91-95.
Link: <https://www.nature.com/articles/s41586-021-04043-8>
4. **MAVERICK** (Zuchner Group, 2023): A supervised model of variant effect for monogenic diseases developed by a CMT research group.
 - **Full citation:** Danzi, M.C., Dohrn, M.F., Fazal, S., Beijer, D., Rebelo, A.P., Cintra, V. and Züchner, S., 2023. Deep structured learning for variant prioritization in Mendelian diseases. *Nature Communications*, 14(1), p.4167.
Link: <https://www.nature.com/articles/s41467-023-39306-7>
5. **Genomic heterogeneity inflates the performance of variant pathogenicity predictions** (Baiyu et al 2025): A recent preprint reporting an extensive benchmark on AI VEP models that shows that the impressive performance scores reported for these models are to be interpreted with care!
 - **Full citation:** Lu, B., Liu, X., Lin, P.Y. and Brandes, N., 2025. Genomic heterogeneity inflates the performance of variant pathogenicity predictions. *bioRxiv*, pp.2025-09.
Link: <https://www.biorxiv.org/content/10.1101/2025.09.05.674459v2>

II. Health data management

Sharing of genetic and other research data is especially critical for progress in rare diseases. The FAIR principles (see below) provide guidelines to ensure the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digitized data. They have been widely endorsed; nevertheless, the effective implementation of these principles into data sharing solutions has proven challenging.

6. **FAIR principles** (Wilkinson et al 2016): The FAIR Guiding Principles for scientific data management and stewardship.
 - **Full citation:** Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), pp.1-9.
 - **Link:** <https://pmc.ncbi.nlm.nih.gov/articles/PMC4792175/>
7. **gnomAD** (Daniel McArthur group, 2020): From ExAC to gnomAD, this database and associated tools remains one of the most critical open-source platforms for genetic data sharing (of allele frequencies in particular).
 - **Full citation:** Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. and Gauthier, L.D., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), pp.434-443.
 - Link:** <https://www.nature.com/articles/s41586-020-2308-7>
 - Website:** <https://gnomad.broadinstitute.org/>
8. **GeneMatcher** (Sobreira et al 2015): Despite it's age, still one of the most important platforms for matching rare variants to disease, through a decentralized global network of biomedical researchers.
 - **Full citation:** Sobreira, N., Schiettecatte, F., Valle, D. and Hamosh, A., 2015. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Human mutation*, 36(10), pp.928-930.
 - Link:** <https://pmc.ncbi.nlm.nih.gov/articles/PMC4833888/>
 - Website:** <https://genematcher.org/>
9. **U.S. Rare Disease Clinical Research Network (RDCRN) and European Reference Network for rare neurological diseases (ERN-RND)**. Somewhat equivalent public bodies that aim at stimulating rare disease research through large scale collaboration including coordinated data collection and sharing. The Inherited Neuropathy Consortium (INC), which has driven much progress in CMT over the last decade, was formed as a RDCRN research consortium The ERN-RND currently lacks a dedicated CMT effort.

- RDCRN: <https://ncats.nih.gov/research/research-activities/rdcrn>
INC: <https://inc.rarediseasesnetwork.org/>
ERN-RND: <https://www.ern-rnd.eu/data-analysis-2023>

10. EURO-NMD (Atalaia et al 2024). An illustrative case study of the challenges of building FAIR data sharing platforms and the complex competing incentives (financial and otherwise) that tend to corrupt these efforts.

- **Full citation:** Atalaia, A., Wandrei, D., Lalout, N., Thompson, R., Tassoni, A., 't Hoen, P.A., Athanasiou, D., Baker, S.A., Sakellariou, P., Paliouras, G. and D'Angelo, C., 2024. EURO-NMD registry: federated FAIR infrastructure, innovative technologies and concepts of a patient-centred registry for rare neuromuscular disorders. *Orphanet journal of rare diseases*, 19(1), p.66.
Link: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10865673/>

We are most grateful for the generous support of this Conference by:



Co-funded by the
European Union

