

# Data-derived wearable digital biomarkers predict *Frataxin* gene expression levels and longitudinal disease progression in Friedreich's Ataxia

Aldo Faisal (✉ [a.faisal@imperial.ac.uk](mailto:a.faisal@imperial.ac.uk))

Imperial College London <https://orcid.org/0000-0003-0813-7207>

**Balasundaram Kadirvelu**

Imperial College London

**Constantinos Gavriel**

Imperial College London

**Sathiji Nageshwaran**

Imperial College London

**Ping Kei Jackson Chan**

Imperial College London

**Stavros Athanasopoulos**

Imperial College London

**Paola Giunti**

University College London

**Valeria Ricotti**

NIHR Great Ormond Street Hospital Biomedical Research Centre/UCL Great Ormond Street Institute of Child Health <https://orcid.org/0000-0003-1800-1987>

**Thomas Voit**

University College London

**Richard Festenstein**

Imperial College

---

## Article

**Keywords:** Friedreich's ataxia, wearable body sensor networks, Frataxin gene, artificial intelligence

**Posted Date:** August 16th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-737100/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Version of Record:** A version of this preprint was published at Nature Medicine on January 19th, 2023.  
See the published version at <https://doi.org/10.1038/s41591-022-02159-6>.

# Abstract

Friedreich's ataxia (FA) is a neurodegenerative disease caused by the epigenetic repression of the *Frx1* gene modulating mitochondrial activity in the brain, which has a diffuse phenotypic impact on patients' motor behavior. Therefore, with current gold-standard clinical scales, it requires 18–24 month-long clinical trials to determine if disease-modifying therapies are at all beneficial. Our high-performance monitoring approach captures the full-movement kinematics from human subjects using wearable body sensor networks from a cohort of FA patients during their regular clinical visits. We then use artificial intelligence to convert these movement data using universal behavior fingerprints into a digital biomarker of disease state. This enables us to predict two different 'gold-standard' clinical scores (SCAFI, SARA) that serve as primary clinical endpoints. Crucially, by performing gene expression analysis on each patient their personal *Frx1* gene expression levels were poorly, if at all, correlated with their clinical scores – fundamentally failing to establish a link between disease mechanism (dysregulated gene expression) and measures to quantify it in the behavioral phenotype. In contrast, our wearable digital biomarker can accurately predict for each patient their personal *FXN* gene expression levels, demonstrating the sensitivity of our approach and the importance of *FXN* levels in FA. Therefore, our data-derived biomarker approach can not only cross-sectionally predict disease and their gene expression levels but also their longitudinal disease trajectory: it is sensitive and accurate enough to detect disease progression with much fewer subjects or shorter time scales than existing primary endpoints. Our work demonstrates that data-derived wearable biomarkers have the potential to substantially reduce clinical trial durations and a first in-human demonstration of reconstructing *FXN* gene expression levels from behavioral data alone.

## Introduction

About 1 in 17 people suffer from rare diseases, and to date 6,000 rare diseases have been identified and new conditions are being described in the medical literature on a regular basis<sup>1,2</sup>. The challenge for drug development and regulatory approval in many rare diseases is their slow progression, and individually small populations. The clinical scales currently used to quantify the progression of neurodegenerative diseases in patients are insensitive to the extremely slow progression of such diseases<sup>3,4</sup>. This means that it might take many months before they reveal any change. Additionally, some of their sub-components have been shown to lack objectivity<sup>5,6,7</sup> as they are mainly based on subjective estimates collected by clinical personnel, often "by eye". Consequently, there is a pressing need for new methodologies that can achieve accurate and objective monitoring of patients' behavior for neurodegenerative assessments (Fig. 1a). This is most important for monitoring progression in clinical trials which are often excessively long as a result. This is because the length of disease-modifying treatment trials is determined by the time it takes for current gold-standard methods of neurological clinical assessments to detect a slowing or stopping of disease progression (Fig. 1b), sometimes taking years (in the case of Friedreich's Ataxia (FA) for 2 years) to conclusively establish disease progression. Moreover, the low sensitivity and accuracy of many current gold-standard "by-eye" assessments, means

that large patient populations are required to sufficiently power those assessments - which in the case of rare diseases typically requires complex multinational clinical trials<sup>8,9</sup>. This makes the development of treatments for rare disease often uneconomical for pharmaceutical and biotechnology companies to pursue, leaving them untreated or addressed through government-backed academically led initiatives.

We use FA as a proof-of-principle model to show how an AI-defined digital biomarker can help to rapidly accelerate and shrink clinical trials for disease-modifying treatments. FA is a triplet-repeat-expansion, neurodegenerative disease with slow progression frequently leading to an early death. It is a rare disease, affecting only 1 in 40,000 Caucasian people<sup>10,11</sup> and there is as yet no known effective disease-modifying treatment. Currently, patient management is aimed at merely maintaining neurobehavioral function for as long as possible<sup>12-14</sup>. Several promising disease-modifying treatment options have been proposed aimed at restoring Frataxin expression<sup>12,15</sup>. The main way of evaluating the progression of FA disease in patients is through a series of score-based behavioral tests evaluated by-eye<sup>4</sup>. The lack of precision biomarkers of progression is due to the complex nature of FA related behavioral change. FA predominantly affects the sensory proprioceptive system causing deterioration in patients' coordination that worsens over time. It typically starts during childhood and the main symptoms include poor balance, gradual loss of strength and position-sense in the limbs, spasticity, the curvature of the spine (kyphoscoliosis), impaired speech, hypertrophic cardiomyopathy and it has also been associated with an increased incidence of diabetes, impairment of bladder control (incontinence), vision and hearing<sup>16,17</sup>. Clinicians observe patients performing various tasks and extract scores, which can semi-quantitatively define the stage of the disease. Amongst the most used tests are the Scale for the Assessment and Rating of Ataxia (SARA)<sup>18</sup> the Composite Cerebellar Functional Severity Score (CCFS)<sup>19</sup> the Inventory of Non-Ataxia Signs (INAS)<sup>20</sup>, Spinocerebellar Ataxia Functional Index (SCAFI)<sup>21</sup> and the Friedreich's Ataxia Rating Scale (FARS)<sup>4</sup>. All these scales can assess patients' motor control and coordination skills through a series of evaluations, such as the 8 Meter Walk (8MW), the 9-Hole Peg Test (9HPT), the finger-nose test, the finger tapping test and the heel-shin-slide test. Evaluating the overall clinical stage of FA disease requires converting patients' physiological information (e.g. stance, walking gait, sitting posture etc.) into numerical scores which in some cases (FARS) includes patient-reported ability to perform Activities of Daily Living (ADL). Therefore, in many cases the process of deriving discrete integer scores results in losing information about the disease state and is subjective.

Though the score-based metrics can quantify the ataxia disease, they are still subject to variability due to the subjective estimates of some of their components<sup>6,22</sup>. This variability can be partially reduced with extensive training of the assessors<sup>23</sup>. These metrics also suffer from low accuracy and sensitivity, resulting in extended periods to measure changes in the behavior in FA and the need for much larger numbers of patients in clinical trials (2 years with 184 patients or 1 year with 548 patients)<sup>9</sup>.

We are at an exciting time in the research of the underlying disease mechanisms in FA which promises to reveal radical novel therapies. FA is caused by a non-coding GAA-triplet repeat expansion within the first intron of the *Frataxin*<sup>24</sup>. This expansion, on both alleles of the gene, leads to partial silencing of the

*Frataxin* gene sufficient to cause Frataxin deficiency and hence the disease. Frataxin protein is essential for normal mitochondrial function and the synthesis of iron-sulfur cluster enzymes, and its deficiency leads to increased susceptibility to oxidative stress. Recent studies by some of us and others have demonstrated that *FXN* gene silencing triggered by the large GAA-repeat expansions involves epigenetic heterochromatinization of the *FXN* gene<sup>25–28</sup>. These findings led to potential disease-modifying therapies aimed at restoring frataxin expression in patients by using epigenetic modifiers such as histone deacetylase inhibitors<sup>12,29,30</sup>. Other promising therapies which act downstream of Frataxin deficiency have also been recently trialed<sup>31</sup> and several large pharmaceutical companies are investing in gene-therapy approaches following recent findings in mouse models<sup>15</sup>.

The lack of precision biomarkers of progression is due to the complex nature of FA related behavioral change. Applying behavior analytics will be crucial in capturing the subtle changes more objectively and accurately and therefore in a shorter timespan than the conventional clinical measures, thereby enabling us to reduce the duration of clinical trials and thus the drug development costs and feasibility of finding a disease-modifying treatment for this rare disease.

Currently, the clinical scales ignore valuable information of how a subject performed the tests (the actual motion trajectories of the patient's limbs) and reduce the assessment to simplistic variables, such as time taken to perform a task. For example, the technique we have employed here, rather than simply timing how long it takes a patient to walk eight meters – captures and analyses ‘how’ they walk, in detail, thereby greatly enriching the depth of the analysis.

## Results

We begin by laying out the structure of our results. Subjects came into the clinic and performed routine activity and clinical scales wearing a motion capture suit. The movement data was analyzed and kinematic features were identified from the suit data. We regressed these kinematic features against the clinical scales and predicted future disease progression. Then we applied the kinematic features to regress against *FXN* gene expression - a key FA molecular biomarker. We compared our features to disease progression with those from a large cohort of clinical scores and show here that our features can outperform the individual's prediction of disease progression by the clinical scales, because we use richer data sources and better features.

Nine FA participants and 3 healthy controls participated in the study. In our study, we used a motion-capture suit to record the behavior of FA patients and healthy controls. All subjects stayed a day and night at our clinical research facility, performing a range of standardized clinical tasks as well as wearing the suit during their daily routine in the clinic - none of the subjects had issues with wearing the suit. To enable monitoring of the FA disease progression on a longitudinal timescale the trial consisted of 4 clinical measurement points: day 1 visit, 3 weeks, 3 months, and 9 months.

To benchmark how FA patients' movement differs from normal we analyzed two sub-assessments of the SCAFI scale: the 8 Meter Walk (8MW) and the 9 Hole Peg Test (9HPT). Currently clinicians only use the crude measure of duration of these activities for quantifying disease severity<sup>21</sup>. We defined a series of kinematic features of patient performance (see Table 1), which can be used to objectively distinguish the differences in the behavior of FA patients from the control population. More specifically, in the 8MW case, we focused on the behavioral changes of the full-body kinematics whereas for the 9HPT we focused on the upper body kinematics since subjects were seated during the task (Fig. 1c-e). The latter is important as current scales exhibit ceiling effects<sup>9</sup> resulting in the exclusion of wheelchair bound patients from clinical trials. These features have been inspired by the currently used clinical scales, standard gait analysis methods, works on other neurodegenerative diseases with similar movement disorders<sup>32-37</sup> and direct clinical experience. Several of these features are intuitively meaningful, perhaps even subjectively visible by eye, while others capture complex and subtle spatiotemporal patterns that may escape even very experienced clinicians. A detailed explanation of all these features is presented in the supplementary methods section.

Table 1

– Suit features from the 8MW task and 9HPT task which were used to train the Gaussian Process regression algorithm against the clinical scales.

ID	Feature name	Description	Num. of features
<b>Suit features from the 8MW task</b>			
F1	Coefficient of variation in the walk cycle duration	Ratio of the standard deviation to the mean of the walk cycle duration	1
F2	Workspace probability density volume & entropy	Volume occupied by the joints calculated using the 3D location of the joints	2
F3	Lower body joint variability	Average variability of the hip and knee joint velocities	6
F4	Walk autocorrelation & decay	Peaks of the three principal components and delay of the autocorrelation of the joint angular velocities	4
F5	Channel-delay cross-correlation	Eigen spectrum values (1, 5 & 35) of the channel-delay cross-correlation matrix	3
F6	Extremities velocity	Average peak velocities of the extremities (wrists and ankles)	4
F7	Walk complexity	Human movement complexity metric and degrees of freedom to explain 90% variance	2
F8	Leg's root mean square power spectrum	Average energy per walk cycle of the hip and knee joint velocities	6
F9	Joint velocities correlation coefficient	Pearson's correlation coefficients between lower body joints	11
F10	Head - spine movement plane area	Area and variability of the head movements on the frontal and sideways plane	3
<b>Suit features from the 9HPT task</b>			
F11	Average upper body joint velocity	Average joint angular velocities of the shoulder and elbow joints	5
F12	Upper body complexity	Human movement complexity metric and degrees of freedom to explain 90% variance of the upper body joint velocities	2
F13	Workspace probability density volume & entropy	Volume occupied by the joints calculated using the 3D location of the joints of the upper body	2

ID	Feature name	Description	Num. of features
F14	Upper body autocorrelation full-width at half-maximum	The width of the autocorrelation curve (of the joint angular velocities of the upper body joints) at the point when it reaches a value of 0.5	5
F15	Channel-delay cross-correlation	Eigen spectrum values (1, 5, 30 & 300) of the channel-delay cross-correlation matrix	4
F16	Arm root mean square power spectrum	Average energy of the shoulder and elbow joint angular velocities	5
F17	Wrist average velocity	Average velocities of the wrist in space	1
F18	Upper body joints' correlations	Pearson's correlation coefficients between upper body joints	3
F19	Logistic fit on upper body joints' velocity	Scale parameter of the logistic distribution of upper body joint's angular velocities and wrist's velocity in space	9
F20	Head spine movement plane area	Area of the head movements on the frontal and sideways plane	1

Most of these features are significantly different between FA patients and controls thereby quantitatively capturing the effects of FA on movement. To develop a clinically useful and improved measure of deterioration we set out to predict the continuous values of the clinical scale quantitatively. We calculated Pearson's correlation coefficient of each feature with respect to the SARA and SCAFI scales and most of the features presented absolute correlations in the range of 0.3–0.5 with respect to the two clinical scales. Therefore, none of them can be independently used for monitoring disease progression. However, a more robust prediction can be potentially achieved by combining all these behavioral features - the same way as is applied in the standard clinical scales.

The relationships linking our features and clinical scales are non-linear. Hence, we used a Gaussian Process (GP) Regression algorithm to find the mapping between the extracted behavioral features and the SARA & SCAFI clinical scales. GP regression is a state-of-the-art method that applies a non-linear regression and can capture the uncertainty in the presence of high variability in the data in a principled manner<sup>38</sup>.

Firstly, we did a cross-sectional prediction of the clinical scales using the suit features from the corresponding visits and the leave one subject out cross validated results are shown in Fig. 2. The algorithm achieved a coefficient of determination ( $R^2$ ) of 0.69 and a root mean square error (RMSE) of 2.54 when predicting SARA scales using suit features of 8MW and a  $R^2$  of 0.45 and a RMSE of 5.10 when using suit features of 9HPT. When predicting SCAFI, the algorithm's performance increased in both the cases of 8MW and 9HPT suit features with  $R^2$  of 0.82 and 0.77, respectively. It should be noted that one

subject could not do the 8MW test and did only the 9HPT. This establishes that our methodology can be used to predict the clinical scales for non-ambulatory patients too. The challenge with leave-one-subject-out cross-validation is that every time the algorithm is tested it is on a new subject with completely new dynamics. Nevertheless, the suit features can still predict the disease state of the patients with good accuracy. The features selected by the feature selection algorithm for predicting SARA and SCAFI are presented in Supplementary Figs. 15 and 16.

Since our patients do not cover the whole range of the clinical scores (0–40), the algorithm's performance is not very good in predicting the SARA scores at the higher end of the scale (e.g., the two sporadic values at the top right corner of the plot). It is clear from the results that the GP regression performance can be improved with a bigger dataset. It can also be observed that both the 8MW and 9HPT suit features are better at predicting SCAFI when compared to SARA. This should not be surprising as both 8MW and 9HPT are part of the SCAFI test suite and the suit features of the 8MW and 9HPT subtasks will have more predictive power at predicting the SCAFI score.

We then wanted to analyze how well the kinematic features extracted from the suit data of 8MW and 9HPT can accurately predict the longitudinal disease progression occurring in FA patients when compared to scores obtained following conventional assessment by clinicians (Fig. 3). First, we wanted to understand how the clinical scales change over a year as a function of their day-1 clinical scale. In Fig. 3a and 3d, we have plotted the change in SARA and SCAFI scales respectively against their day-1 clinical scale for the FA patients of our study and also patients from EFACTS study<sup>9</sup> (a two-year longitudinal study with a larger cohort size).

We used GP regression to predict the month-9 SARA and SCAFI scales of the subjects from our study using the suit features from day-1 8MW and suit features from day-1 9HPT. We also predicted the month-12 SARA and SCAFI scales of the EFACTS study using the day-1 SARA and SCAFI scores from the EFACTS study. The results of the longitudinal prediction for the SARA and SCAFI are presented in Fig. 3b and Fig. 3e respectively. For the longitudinal predictions of SARA, both the day-1 suit features of 8MW and 9HPT achieved a good leave-one-subject-out cross-validated RMSE of 1.70 and 1.16 in comparison with a RMSE of 2.31 using day-1 SARA. Again, for the longitudinal predictions of SCAFI, the 8MW and 9HPT day-1 suit features outperformed day-1 SCAFI (RMSEs of 0.09 and 0.11 vs 0.25). Please see Supplementary Fig. 17 for plot of the  $R^2$  of the results. This implies that our suit features contained sufficiently rich information not only to score the disease state of the patient in the present, but also predict how the disease would evolve. The features selected by the feature selection algorithm for the longitudinal predictions of SARA and SCAFI are presented in Supplementary Figs. 18 and 19.

We then plotted the RMSE of the longitudinal predictions as a function of number of subjects used to build the machine learning model (Fig. 3c and Fig. 3e). The model using suit features achieved better performance with a smaller number of subjects ( $n = 7$ ) compared to the model using the clinical scales with a larger cohort from the EFACTS study ( $n = 425$ ). This establishes that a small population size is

sufficient to build prediction models with high accuracy when using the rich set of suit features which would therefore significantly reduce the numbers of patients required in the context of drug development.

FA is caused by a GAA-repeat expansion in the *FXN* gene leading to transcriptional repression of *FXN* and the disease. The length of GAA-repeat has been shown to correlate with age of onset and inversely correlate with disease severity. We therefore hypothesized that *FXN* blood levels might inversely correlate with disease severity. We predicted the *FXN* mRNA levels of the subjects using 4 sets of predictors: 8MW suit features, 9HPT suit features, SARA and SCAFI and the results for the leave-one-subject-out cross-validation is presented in Fig. 4a-d. The scatterplot of the measured vs predicted *FXN* levels using are presented in the first two columns. The first column shows the predictions using the 8MW & 9HPT suit features and the second column shows the predictions using the SARA and SCAFI clinical scales. 8MW and 9HPT achieved a  $R^2$  of 0.60 and 0.53 (and a RMSE of 0.53 and 0.62) for the leave-one-subject-out cross-validated case. In comparison, both SARA and SCAFI achieved only  $R^2$  values close to zero (with RMSE values of 0.89 and 1.00). (Please see Supplementary Fig. 20a and 20b for the features selected by the feature selection algorithm for the prediction of *FXN* and Supplementary Fig. 20c for a scatter plot of the *FXN* against SARA and SCAFI).

The total scores of SARA and SCAFI, might be poorer at prediction because they contain less information. We reasoned that by using the individual components of the SARA and SCAFI scales as predictors, we would improve the predictive capacity (see Supplementary Fig. 21). Although this led to an improvement ( $R^2$  of predictions using the components of SARA increased to 0.33 and that of SCAFI increased to 0.19) the prediction using the suit features of the 8MW ( $R^2$  of 0.60) and 9HPT ( $R^2$  of 0.53) still outperforms the individual components of SARA and SCAFI in predicting *FXN* levels.

## Discussion

Clinical scales which attempt to quantify the clinical examination are widely used as gold-standard neurological and cardiovascular assessments are often measured “by-eye” and subject to intra-rater variability (the same assessors reporting different outcomes on repeated trials)<sup>7</sup> and inter-rater variability (different assessors scoring the same assessment differently)<sup>39</sup> as well as variability in the patient performance ability (which may vary based on time and days of the week, seasons etc.). Such clinical scales are composed of a battery of different behavioral sub-assessments (e.g., walking, picking up things) to capture the performance across different domains of the body’s capabilities and average out variability across individual sub-assessments (e.g., for FA SCAFI has 3 sub-assessments while SARA has 8 sub-assessments). This results in longer times to perform clinical tests and more opportunity for rater variability. Moreover, variability from each sub-assessment may, in the worst case, be additive.

We show here that the kinematic features from one test can predict the clinical scales of another test, future scores and indeed the key molecular marker - *FXN* levels, indicating that they carry crucial information about the disease and can directly link behavior with *FXN* gene expression.

Our data-driven approach also revealed new observational information that can complement the current clinical evaluations. For example, one of our main findings, that was supported by multiple metrics, was the significantly reduced movement structure and complexity in the patients' activities compared to controls during the two subtests. The high accuracy of the collected data can also be used to better understand the subtle changes in behavior secondary to the disease or ameliorated by therapeutic intervention. Such increased accuracy would effectively reduce both the duration and cost of clinical trials but also potentially enable detection of this disease and others at very early stages, even before they can be visually observed.

The fact that the kinematic features of each of the subtasks (8MW and 9HPT) can predict the overall clinical scores with good accuracy suggests that our data-driven approach can predict the overall performance of a patient with a minimal number of tasks in the clinical assessment and therefore with much less patient effort. Our data-driven analysis also allowed us to characterize the kinematic alterations in FA patients and verify and objectify the various observations that clinicians use in their verbal characterization such as slower irregular movements.

Because the SARA scale has been shown to have a ceiling effect, FA patients are frequently excluded from clinical trials above a certain SARA score threshold. The fact that we can accurately predict SARA scores using an upper limb coordination test (9HPT), suggests that this might have an impact on those patients that can be included in clinical trials if this novel methodology is validated in a larger study.

Recent efforts<sup>40-44</sup> used movement sensors in an attempt to overcome the errors caused by subjective clinical scales. Parkinson's disease patients<sup>45</sup> donned full-body suit to capture the patients' behavior during clinical tasks. The analysis on cross-sectional classification of the severity of the disease (mild vs severe) rather than capturing changes in the disease progression through time. Similarly, a recent cross-sectional study used wearable motion sensors to quantify the SARA scale and thereby identify very early signs in the spinocerebellar ataxias<sup>46</sup>. Here we go further in developing algorithms to accurately predict outcome longitudinally and moreover predict a molecular biomarker of FA.

This work promises to not only shorten clinical trials that would otherwise be prohibitively long, but also to reduce the number of patients required to measure the deterioration of the neurological state. The latter is especially important in rare neurodegenerative diseases. Therefore, such technology is likely to greatly facilitate the possibility of finding treatment for these relatively neglected and incurable conditions.

We show here that FA provides a model neurodegenerative condition where a holistic machine learning analysis of full-body kinematics from a longitudinal study demonstrates accurate prediction of dysfunction progression in individual patients. Crucially, we not only determine clinical phenotype but also track the molecular cause of the disease (repressed frataxin) from movement data alone. While high-resolution behavioral genomics was previously successful in genetic model organisms such as *C.Elegans*<sup>46</sup> at differentiating mutants through pure digital behavioral analysis and was able to pick up subtle changes in physiological and reproductive state from the movement behavior of *Drosophila*

*Melanogaster*<sup>A7</sup> such data-driven approaches have been lacking in human clinical and genomic applications. Existing digital biomarkers applied in human clinical settings involve supplanting conventional measures such as distance walked on a treadmill with digital measured proxies obtained through digital devices (such as step counters). However, these approaches overlook the richness of signal contained in full-body kinematic data and instead revert to re-using existing clinical measures often defined over 100 years ago. Such data-derived behavior promises to greatly shorten the length or number of patients required for clinical trials in FA and other slowly progressive genetic disorders. To date, the monitoring of neurodegenerative disease progression frequently fails to adequately test disease-modifying therapies because it is slow and of low precision making drug development risky and expensive. Our digital behavioral biomarker approach could flip many rare diseases for which drug development was slow and often too expensive, into viable candidates for rapid treatment advances.

## Declarations

**Acknowledgements:** This study was supported by the UK MRC, Imperial College NIHR BRC, UCL BRC and Ataxia UK. The study was conducted in the NIHR Imperial College Clinical Research Facility. Ethical approval was obtained from the Riverside Ethics Committee (11/LO/0998). We thank all study participants for their commitment to the study and for attending the hospital visits. We thank the EFACTS consortium for sharing the EFACTS study data.

**Author Contributions:** AAF and RF conceived and supervised the conduct of the study. RF was responsible for the initial screening of the subjects. PG assessed and provided suitable patients for deep phenotyping. CG, SN, SA and RF acquired the data. JC performed the FXN analysis. AAF and RF conceived the ethomic biomarker concept and approach. CG, BK and AAF performed the analyses of the data, and all authors contributed to the interpretation for the global data set. VR and TV interpreted the clinical data generated. RF and AAF obtained funding. AAF, BK, RF wrote the draft of the manuscript, and all authors contributed or commented on its revision.

### Competing interests:

AAF has consulted in the domain of AI for Airbus and Celestial Group. He received research funding paid to his institution from Fresenius-KABI, Shell, Airbus. He received donations given to his institution from Facebook, NVIDIA and Microsoft.

RF has consulted for Pfizer and has received funding from Pfizer but not related to this study.

VR was a Solid Biosciences employee at the time of the conduct of the study. Currently VR is Co-Founder EVP, CMO and stockholder of DiNAQOR AG, the products of which are not related to this study.

TV was a consultant for Antisense Therapeutics, BioPhytis, Catabasis, DZHK, Italfarmaco, Santhera, Sarepta, Servier, Syneos Health, Solid Biosciences. TV is an SAB member and stockholder of Constant Pharmaceuticals and a Co-Founder, CSO and stockholder of DiNAQOR AG.

AF, RF, TV, and VR are co-inventors of an International Patent Application PCT/EP2020/077703 - 'Systems and methods for monitoring the state of a disease using a biomarker'.

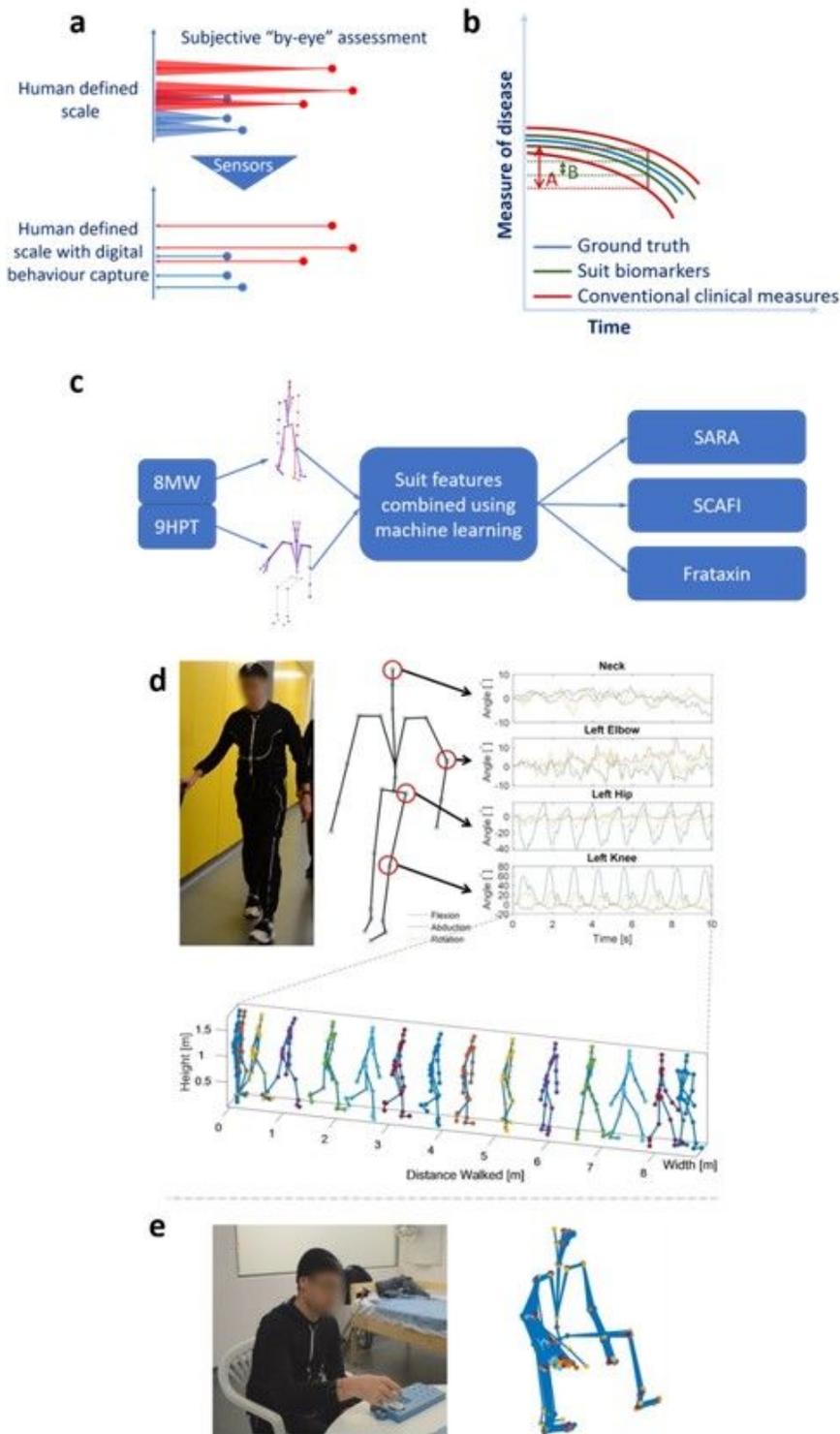
## References

1. European Commission – European Commission. (2020). Rare diseases. [https://ec.europa.eu/info/research-and-innovation/research-area/health/rare-diseases\\_en](https://ec.europa.eu/info/research-and-innovation/research-area/health/rare-diseases_en)
2. Wakap, S.N., *et al.* Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics* **28**, 165–173 (2020).
3. Mancini, M., *et al.* Postural sway as a marker of progression in Parkinson's disease: a pilot longitudinal study. *Gait & posture* **36**, 471–476 (2012).
4. Fahey, M.C., Corben, L., Collins, V., Churchyard, A.J. & Delatycki, M.B. How is disease progress in Friedreich's ataxia best measured? A study of four rating scales. *J Neurol Neurosurg Psychiatry* **78**, 411–413 (2007).
5. Albanese, M.A., Clarke, W.R., Adams Jr, H.P. & Woolson, R.F. Ensuring reliability of outcome measures in multicenter clinical trials of treatments for acute ischemic stroke. The program developed for the Trial of Org 10172 in Acute Stroke Treatment (TOAST). *Stroke* **25**, 1746–1751 (1994).
6. Subramony, S., *et al.* Measuring Friedreich ataxia: interrater reliability of a neurologic rating scale. *Neurology* **64**, 1261–1262 (2005).
7. Burk, K., Schulz, S.R. & Schulz, J.B. Monitoring progression in Friedreich ataxia (FRDA): the use of clinical scales. *J Neurochem* **126 Suppl 1**, 118–124 (2013).
8. Reetz, K., *et al.* Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: a cross-sectional analysis of baseline data. *The Lancet Neurology* **14**, 174–182 (2015).
9. Reetz, K., *et al.* Progression characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS): a 2 year cohort study. *The Lancet Neurology* **15**, 1346–1354 (2016).
10. Koeppen, A.H. Friedreich's ataxia: pathology, pathogenesis, and molecular genetics. *Journal of the neurological sciences* **303**, 1–12 (2011).
11. Koeppen, A.H. Nikolaus Friedreich and degenerative atrophy of the dorsal columns of the spinal cord. *Journal of neurochemistry* **126**, 4–10 (2013).
12. Libri, V., *et al.* Epigenetic and neurological effects and safety of high-dose nicotinamide in patients with Friedreich's ataxia: an exploratory, open-label, dose-escalation study. *The Lancet* **384**, 504–513 (2014).
13. Lodi, R., *et al.* Antioxidant treatment improves in vivo cardiac and skeletal muscle bioenergetics in patients with Friedreich's ataxia. *Annals of neurology* **49**, 590–596 (2001).
14. Lynch, D.R. & Fischbeck, K.H. Nicotinamide in Friedreich's ataxia: useful or not? *The Lancet* **384**, 474–475 (2014).

15. Ocana-Santero, G., Diaz-Nido, J. & Herranz-Martin, S. Future Prospects of Gene Therapy for Friedreich's Ataxia. *Int J Mol Sci* **22**(2021).
16. Durr, A., *et al.* Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N Engl J Med* **335**, 1169–1175 (1996).
17. Gibilisco, P. & Vogel, A.P. Friedreich ataxia. *BMJ* **347**, f7062 (2013).
18. Schmitz-Hübsch, T., *et al.* Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* **66**, 1717–1720 (2006).
19. Du Montcel, S.T., *et al.* Composite cerebellar functional severity score: validation of a quantitative score of cerebellar impairment. *Brain* **131**, 1352–1361 (2008).
20. Jacobi, H., *et al.* Inventory of Non-Ataxia Signs (INAS): validation of a new clinical assessment instrument. *The Cerebellum* **12**, 418–428 (2013).
21. Schmitz-Hübsch, T., *et al.* SCA Functional Index: a useful compound performance measure for spinocerebellar ataxia. *Neurology* **71**, 486–492 (2008).
22. Fillyaw, M.J., *et al.* Quantitative measures of neurological function in chronic neuromuscular diseases and ataxia. *Journal of the neurological sciences* **92**, 17–36 (1989).
23. Storey, E., Tuck, K., Hester, R., Hughes, A. & Churchyard, A. Inter-rater reliability of the international cooperative Ataxia Rating Scale (ICARS). *Movement disorders: official journal of the Movement Disorder Society* **19**, 190–192 (2004).
24. Campuzano, V., *et al.* Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**, 1423–1427 (1996).
25. Chan, P.K., *et al.* Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. *Hum Mol Genet* **22**, 2662–2675 (2013).
26. Chutake, Y.K., Lam, C., Costello, W.N., Anderson, M. & Bidichandani, S.I. Epigenetic promoter silencing in Friedreich ataxia is dependent on repeat length. *Ann Neurol* **76**, 522–528 (2014).
27. De Biase, I., Chutake, Y.K., Rindler, P.M. & Bidichandani, S.I. Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. *PLoS One* **4**, e7914 (2009).
28. Saveliev, A., Everett, C., Sharpe, T., Webster, Z. & Festenstein, R. DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature* **422**, 909–913 (2003).
29. Festenstein, R. Breaking the silence in Friedreich's ataxia. *Nat Chem Biol* **2**, 512–513 (2006).
30. Gottesfeld, J.M. Small molecules affecting transcription in Friedreich ataxia. *Pharmacol Ther* **116**, 236–248 (2007).
31. Lynch, D.R., *et al.* Safety and Efficacy of Omaveloxolone in Friedreich Ataxia (MOXIe Study). *Annals of Neurology* **89**, 212–225 (2021).
32. Hollman, J.H., Kovash, F.M., Kubik, J.J. & Linbo, R.A. Age-related differences in spatiotemporal markers of gait stability during dual task walking. *Gait & posture* **26**, 113–119 (2007).

33. Plotnik, M., Giladi, N. & Hausdorff, J.M. Bilateral coordination of walking and freezing of gait in Parkinson's disease. *European Journal of Neuroscience* **27**, 1999–2006 (2008).
34. Sofuwa, O., *et al.* Quantitative gait analysis in Parkinson's disease: comparison with a healthy control group. *Archives of physical medicine and rehabilitation* **86**, 1007–1013 (2005).
35. Van Meulen, F.B., *et al.* Objective evaluation of the quality of movement in daily life after stroke. *Frontiers in bioengineering and biotechnology* **3**, 210 (2016).
36. van Meulen, F.B., Reenalda, J. & Veltink, P.H. Assessing of motor performance in stroke using body worn sensing. in *Replace, Repair, Restore, Relieve—Bridging Clinical and Engineering Solutions in Neurorehabilitation* 191–193 (Springer, 2014).
37. Yu, L., Xiong, D., Guo, L. & Wang, J. A remote quantitative Fugl-Meyer assessment framework for stroke patients based on wearable sensor networks. *Computer methods and programs in biomedicine* **128**, 100–110 (2016).
38. Rasmussen, A., Gomez, M., Alonso, E. & Bidichandani, S.I. Clinical heterogeneity of recessive ataxia in the Mexican population. *J Neurol Neurosurg Psychiatry* **77**, 1370–1372 (2006).
39. Hobson, J., Edwards, N. & Meara, R. The Parkinson's Disease Activities of Daily Living Scale: a new simple and brief subjective measure of disability in Parkinson's disease. *Clinical Rehabilitation* **15**, 241–246 (2001).
40. Jobbágy, Á., Harcos, P., Karoly, R. & Fazekas, G. Analysis of finger-tapping movement. *Journal of neuroscience methods* **141**, 29–39 (2005).
41. McPartland, M.D., Krebs, D.E. & Wall, C. Quantifying ataxia: ideal trajectory analysis—a technical note. *Journal of rehabilitation research and development* **37**, 445–454 (2000).
42. Zampieri, C., *et al.* The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry* **81**, 171–176 (2010).
43. Gavriel C, T.A., Lourenco PR, Nageshwaran S, Athanasopoulos S, Sylaidi A, Festenstein R, Faisal A. Kinematic body sensor networks and behaviourmetrics for objective efficacy measurements in neurodegenerative disease drug trials. *IEEE Body Sensor Networks Conference* 2015 (2015).
44. Gavriel C, T.A., Lourenco P, Nageshwaran S, Athanasopoulos S, Sylaidi A, Festenstein R, Faisal AA. Towards neurobehavioral biomarkers for longitudinal monitoring of neurodegeneration with wearable body sensor networks. *IEEE Neural Engineering (NER)*, 348–351 (2015).
45. Das, S., *et al.* Quantitative measurement of motor symptoms in Parkinson's disease: A study with full-body motion capture data. in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 6789–6792 (IEEE, 2011).
46. Kashyap, B., *et al.* objective Assessment of cerebellar Ataxia: A comprehensive and Refined Approach. *Scientific reports* **10**, 1–17 (2020).

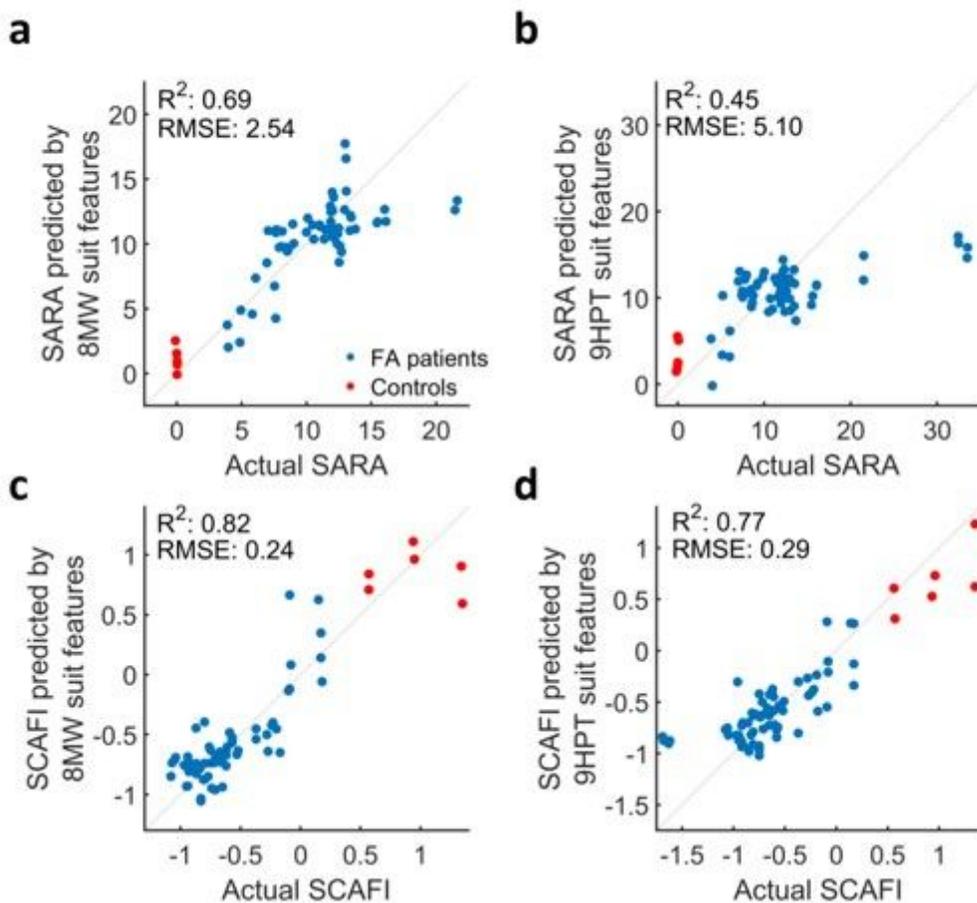
## Figures



**Figure 1**

Overview of concepts, methodology and findings: a. Current gold-standard clinical assessments for monitoring the effects of neurodegenerative diseases are often measured "by-eye" and lack objectivity. Sensors can help achieve accurate and objective monitoring of patients' behavior. b. Because of the lack of precision, it takes months before conventional clinical measures reveal any change in the disease progression. Motion-capture suit biomarkers are more accurate and sensitive providing digital biomarkers

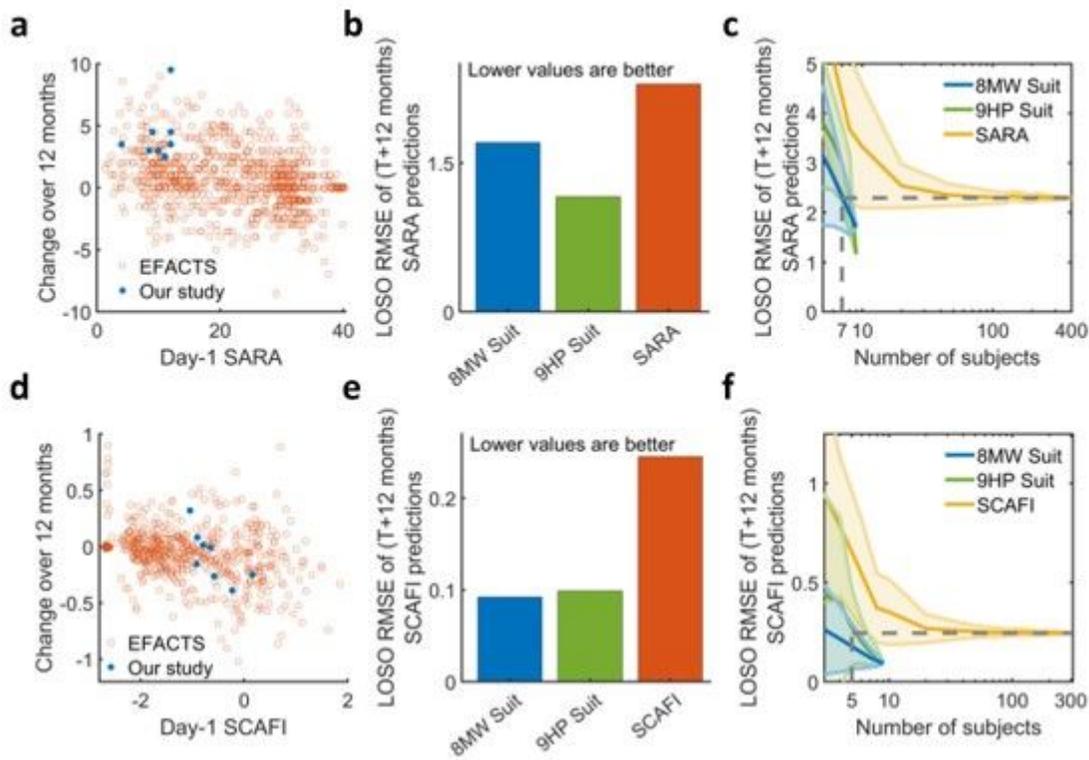
that can detect even small changes in patient performance. Applying behavior analytics can capture the subtle changes more objectively and in a shorter time span than the conventional clinical measures. c. Using a full-body motion capture approach, we analyzed two sub-assessments of the SCAFI scale, the 8 Meter Walk (8MW) and the 9 Hole Peg Test (9HPT), for which clinicians only use their duration for estimating the progression of the FA disease. We applied a machine learning approach and generated a series of markers of patient performance. Using this data-driven approach we reconstructed the full SARA and SCAFI scores along with the Frataxin levels. d. A subject performing the 8MW while wearing the IGS-180 motion capture suit [Animazoo UK Ltd, Brighton, UK], which uses 17 inertial sensors to monitor the movement of the limbs and a reconstruction of the body posture at a single frame. A typical time-series from the angular positions of the neck, elbow, hip and knee joints captured by the suit and a time-lapse of an FA patient performing the 8MW test (each frame is 0.5s apart). e. The setup used during the 9HPT showing the subject with the motion capture suit and a time-lapse of an FA patient performing the 9HPT.



**Figure 2**

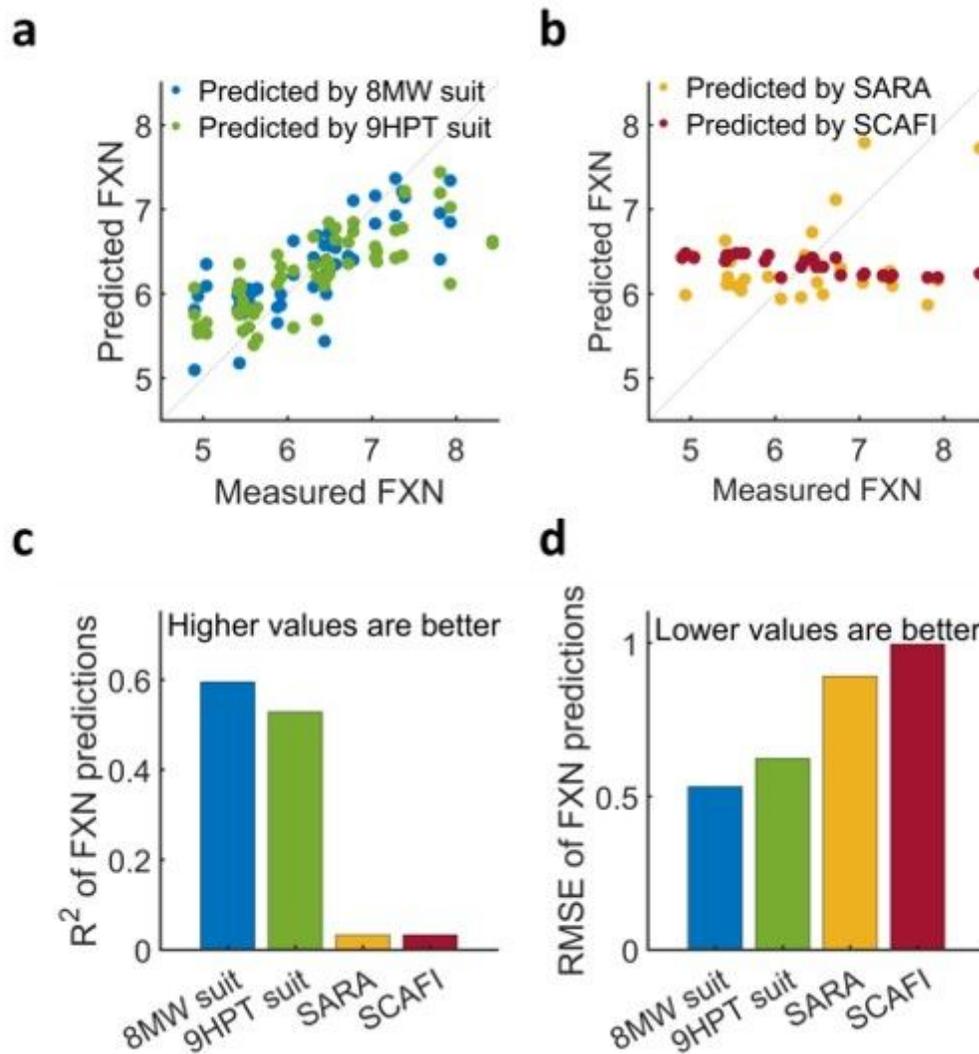
Cross-sectional predictions of SARA and SCAFI: Actual vs predicted plot from the cross-sectional prediction of the clinical scales using the suit features from the 8MW and 9HPT tasks. Each point represents a single actual vs predicted score of an FA Patient (9 subjects, 33 visits) or Control (3 subjects, 3 visits). One subject could not do the 8MW and did only the 9HPT. There are two points per clinical visit of a subject as there are two suit measurements for 8MW and 9HPT from each visit. A small jitter is

added to the points to show overlapping points. A leave-one-subject-out cross-validation and Gaussian Process regression was used to find a mapping between the suit features and the clinical scales .



**Figure 3**

Longitudinal predictions of SARA and SCAFI: a. Scatter plot of the day-1 SARA vs the change in the SARA scale over 9 months for data from our study ( $n=8$ ) and change in the SARA scale over 12 months for the EFACTS study ( $n=425$ ). A small jitter is added to the points to show overlapping points. b. Comparison of the aggregate RMSE of the leave-one-subject-out cross-validated predictions of the SARA at T+ 9 months by the suit features of 8MW and 9HPT tasks from visit T of our study and SARA at T+ 12 months predicted using SARA at visit T from EFACTS study c. Plot of the aggregate leave-one-subject-out cross-validated RMSE of the predictions of the SARA at T+ 9 months from our study (T+12 months in case of EFACTS) as a function of number of subjects used to build the machine learning model. The model using suit features achieves better performance with a smaller number of subjects (marked by the vertical dotted lines) compared to the model using SARA with a much larger cohort from the EFACTS study. The shaded regions indicate the standard deviation of the results from different models built for each number of subjects. d-e are the corresponding plots for the SCAFI scale.



**Figure 4**

Cross sectional predictions of FXN gene expression: FXN mRNA levels were predicted using the following 4 sets of predictors: suit features from the 8MW task, suit features from the 9HPT task, SARA and SCAFI scores. a. Scatter plot of the measured FXN mRNA levels against the FXN mRNA levels predicted by the suit features from the 8MW and 9HPT tasks of the corresponding visits. Please note that there are two points per clinical visit of a subject for the suit features as there are two suit measurements for 8MW and 9HPT from each visit. b. Scatter plot of the measured FXN mRNA levels against the FXN mRNA levels predicted by the SARA and SCAFI clinical scales of the corresponding visits. FXN mRNA was measured by Q-RT-PCR and the CT values normalised to TBP are plotted. c & d. Leave-one-subject-out cross-validated R<sup>2</sup> and RMSE of the predictions by the 4 set of predictors. GP regression models using the features of the suit data of 8MW and 9HPT tasks as predictors perform better compared to GP regression models using the SARA and SCAFI scales as predictors.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Methods.docx](#)
- [SupplementaryTablesAndFigures.docx](#)