

RESEARCH ARTICLE

# OCR-Stats: Robust estimation and statistical testing of mitochondrial respiration activities using Seahorse XF Analyzer

Vicente A. Yépez<sup>1,2</sup>, Laura S. Kremer<sup>3,4</sup>, Arcangela Iuso<sup>3,4</sup>, Mirjana Gusic<sup>3,4</sup>, Robert Kopajtich<sup>3,4</sup>, Eliška Koňářiková<sup>3,4</sup>, Agnieszka Nadel<sup>3,4</sup>, Leonhard Wachutka<sup>1</sup>, Holger Prokisch<sup>3,4</sup>, Julien Gagneur<sup>1,2\*</sup>

**1** Department of Informatics, Technical University of Munich, Garching, Germany, **2** Quantitative Biosciences Munich, Gene Center, Department of Biochemistry, Ludwig-Maximilians Universität München, Munich, Germany, **3** Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany, **4** Institute of Human Genetics, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

\* [gagneur@in.tum.de](mailto:gagneur@in.tum.de)



**OPEN ACCESS**

**Citation:** Yépez VA, Kremer LS, Iuso A, Gusic M, Kopajtich R, Koňářiková E, et al. (2018) OCR-Stats: Robust estimation and statistical testing of mitochondrial respiration activities using Seahorse XF Analyzer. PLoS ONE 13(7): e0199938. <https://doi.org/10.1371/journal.pone.0199938>

**Editor:** Jianhua Zhang, University of Alabama at Birmingham, UNITED STATES

**Received:** March 9, 2018

**Accepted:** June 16, 2018

**Published:** July 11, 2018

**Copyright:** © 2018 Yépez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This study was supported by the German Bundesministerium für Bildung und Forschung (BMBF) through the German Network for mitochondrial disorders (mitoNET, 01GM1113C to HP), the E-Rare project GENOMIT (01GM1207 to HP), the Juniorverbund in der Systemmedizin 'mitOmics' (FKZ 01ZX1405A JG, LW and VAY), the DZHK (German Centre for Cardiovascular

## Abstract

The accurate quantification of cellular and mitochondrial bioenergetic activity is of great interest in medicine and biology. Mitochondrial stress tests performed with Seahorse Bioscience XF Analyzers allow the estimation of different bioenergetic measures by monitoring the oxygen consumption rates (OCR) of living cells in multi-well plates. However, studies of the statistical best practices for determining aggregated OCR measurements and comparisons have been lacking. Therefore, to understand how OCR behaves across different biological samples, wells, and plates, we performed mitochondrial stress tests in 126 96-well plates involving 203 fibroblast cell lines. We show that the noise of OCR is multiplicative, that outlier data points can concern individual measurements or all measurements of a well, and that the inter-plate variation is greater than the intra-plate variation. Based on these insights, we developed a novel statistical method, OCR-Stats, that: i) robustly estimates OCR levels modeling multiplicative noise and automatically identifying outlier data points and outlier wells; and ii) performs statistical testing between samples, taking into account the different magnitudes of the between- and within-plate variations. This led to a significant reduction of the coefficient of variation across plates of basal respiration by 45% and of maximal respiration by 29%. Moreover, using positive and negative controls, we show that our statistical test outperforms the existing methods, which suffer from an excess of either false positives (within-plate methods), or false negatives (between-plate methods). Altogether, this study provides statistical good practices to support experimentalists in designing, analyzing, testing, and reporting the results of mitochondrial stress tests using this high throughput platform.

Research, LSK), the German Research Foundation (DFG) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program. A Fellowship through the Graduate School of Quantitative Biosciences Munich (QBM) supports VAY. HP is supported by EU FP7 Mitochondrial European Educational Training Project (317433). JG, VAY, LSK and RK and HP are supported by EU Horizon2020 Collaborative Research Project SOUND (633974). We thank the Cell lines and DNA Bank of Pediatric Movement Disorders and Mitochondrial Diseases of the Telethon Genetic Biobank Network (GTB09003).

**Competing interests:** The authors have declared that no competing interests exist.

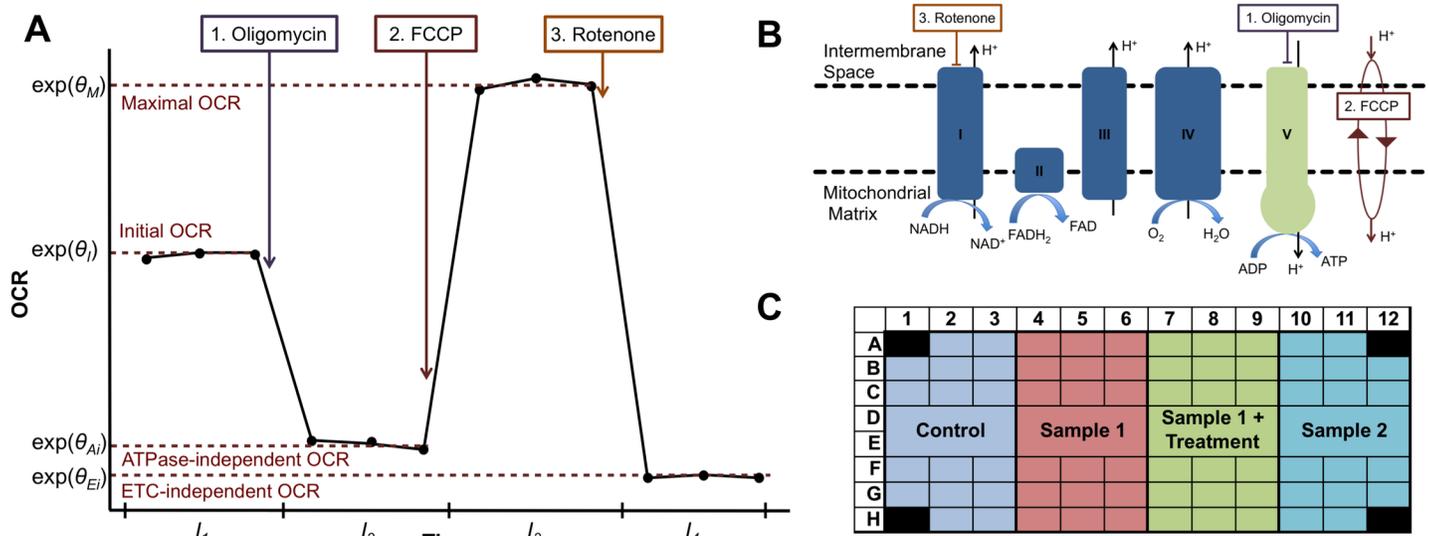
## Introduction

Mitochondria are double-membrane-enclosed, ubiquitous, maternally inherited organelles present in most eukaryotic cells [1]. They are known as the powerhouse of the cell [2,3] due to their pivotal function in the cellular energy supply where adenosine triphosphate (ATP) is generated by the mitochondrial respiratory chain in a process referred to as oxidative phosphorylation. Furthermore, mitochondria are involved in regulating reactive oxygen species [4], apoptosis [2], amino acid synthesis [5,6], cell proliferation [6], cell signaling [7], and in the regulation of innate and adaptive immunity [8]. A decline in mitochondrial function, reflected by a diminished electron transport chain activity, is related to many human diseases ranging from rare genetic disorders [9] to common ones such as cancer [7,10], diabetes [11], neurodegeneration [12], and aging [3]. One of the most informative tests of mitochondrial function is the quantification of cellular respiration, since it directly reflects electron transport chain impairment [9] and depends on many sequential reactions from glycolysis to oxidative phosphorylation [13]. One of the last steps of cellular respiration is the oxidation of cytochrome c in complex IV, which reduces oxygen to form water. Therefore, the estimations of oxygen consumption rates (OCR) expressed in pmol/min enable drawing conclusions about the ability to synthesize ATP and about mitochondrial function, even more than the measurements of intermediates (such as ATP or nicotinamide adenine dinucleotide NADH) and potentials [14,15].

OCR was classically measured using a Clark-type electrode, which is time-consuming, limited to whole cells in suspension and high yield, and does not allow the automated injection of compounds [15]. In addition, experimentation with isolated mitochondria is ineffective because the cellular regulation of mitochondrial function is removed during isolation [16]. In the last few years, a new technology that calculates O<sub>2</sub> concentrations from fluorescence [17] in a microplate assay format has been developed by the company Seahorse Bioscience (now part of Agilent Technologies) [18]. It allows simultaneous real-time measurements of both OCR and extracellular acidification rate (ECAR) in multiple cell lines and conditions, reducing the amount of required sample material and increasing the throughput [18,19].

Typically, OCR and ECAR are measured using the Seahorse XF Analyzer in 96-well (or 24-well) plates at multiple time steps under three consecutive treatments (Fig 1), in a procedure known as a mitochondrial stress test [20]. Under basal conditions, complexes I–IV exploit energy derived from electron transport to pump protons across the inner mitochondrial membrane. The proton gradient generated in this manner is subsequently harnessed by complex V to generate ATP. The blockage of the proton translocation through complex V by injecting oligomycin represses ATP production and prevents the electron transport throughout complexes I–IV due to the unexploited gradient, thus, generating ATP-ase independent OCR only (Fig 1A and 1B). The administration of carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP), an ionophor, subsequently dissipates the gradient uncoupling electron transport from complex V activity and increasing oxygen consumption to a maximum level (Fig 1A and 1B). Finally, mitochondrial respiration is completely halted using rotenone, a complex I inhibitor. There is still some remaining oxygen consumption that is independent from electron transport chain activity (Fig 1A and 1B). This approach is label-free and non-destructive, so the cells can be retained and used for further assays [21].

OCR differences in the natural scale between the various stages of this procedure lead to the estimation of six different bioenergetic measures: basal respiration, proton leak, non-mitochondrial respiration, ATP-linked respiration, spare respiratory capacity, and maximal respiration [15,19] (Table 1). An increase in proton leak and a decrease in basal or maximal respiration are indicators of mitochondrial dysfunction [15]. In addition, ATP-linked respiration, basal respiration, and spare capacity change in response to ATP demand, which is not



**Fig 1. Principle of the mitochondrial stress test assay.** (A) Cartoon illustration of OCR levels (y-axis) versus time (x-axis). Injection of the three compounds oligomycin, FCCP, and rotenone delimits four time intervals within each of which OCR is roughly constant. (B) Targets of each compound in the electron transport chain. (C) Typical layout of a mitochondrial stress test 96-well plate.

<https://doi.org/10.1371/journal.pone.0199938.g001>

necessarily mitochondrion-related as it may be the consequence of the dysregulation of any cellular process altering general cellular energy demand. Then, these bioenergetic measures are typically used to test two samples or conditions against each other.

The existing literature describing the Seahorse technology addresses experimental aspects regarding sample preparation [22,23], the number of cells to seed [23,24], and compound concentration in different organisms [13,22,25]. However, studies regarding statistical best practices for determining OCR levels and testing them against others are lacking. The sole definition of bioenergetic measures varies between authors, as well as the number of time points in each interval (usually three time points, but in some cases one [26], two [27], or four or more [11]), and whether differences [6,13,28], ratios [12,29], or both [24,25] should be

**Table 1. OCR ratios, abbreviations, definitions, metrics, and analogous definitions.**

OCR ratios	Abbr.	Definition	Metrics	Analogous in literature
ETC-dependent OCR proportion	E/I-proportion	Proportion of OCR in the ETC with respect to the initial OCR	$\frac{OCR_1 - OCR_4}{OCR_1} = 1 - \exp(\theta_{Ei} - \theta_1)$	Basal respiration: $OCR_1 - OCR_4$
ATPase-dependent OCR proportion	A/I-proportion	Proportion of OCR driven from ATPase proton pumping with respect to the initial OCR	$\frac{OCR_1 - OCR_2}{OCR_1} = 1 - \exp(\theta_{Ai} - \theta_1)$	ATP-linked respiration: $OCR_1 - OCR_2$
ETC-dependent proportion of ATPase-independent OCR	E/Ai-proportion	Proportion of OCR in the ETC, but not driven from ATPase proton pumping, with respect to all non ATPase driven OCR	$\frac{OCR_2 - OCR_4}{OCR_2} = 1 - \exp(\theta_{Ei} - \theta_{Ai})$	Proton leak: $OCR_2 - OCR_4$
Maximal over initial OCR fold change	M/I-fold change	Ratio between maximal OCR and initial OCR	$\frac{OCR_3}{OCR_1} = \exp(\theta_M - \theta_1)$	Spare respiratory capacity: $OCR_3 - OCR_1$
Maximal over ETC-independent OCR fold change	M/Ei-fold change	Ratio between maximal OCR and non-ETC driven OCR	$\frac{OCR_3}{OCR_4} = \exp(\theta_M - \theta_{Ei})$	Maximal respiration: $OCR_3 - OCR_4$
Not defined as a ratio	NA	NA	NA	Non-mitochondrial respiration: $OCR_4$

Proposed definitions for cellular bioenergetics based on ratios, their abbreviations, equations to compute them and analogous measures used in the literature.  $OCR_i$  and  $\theta_i$  correspond to the expected OCR value on time interval  $i$  in the natural and logarithmic scale, respectively (Fig 1A).

<https://doi.org/10.1371/journal.pone.0199938.t001>

computed. Consequently, the comparison of results across studies is difficult. Moreover, often, statistical power analyses for experimental design are not provided. The differences in OCR between biological samples (e.g. patient vs. control, or gene knockout vs. WT) can be as low as 12%–30% [30–32]. Therefore, to design experiments with appropriate power to significantly detect such differences, it is important to know the source and amplitude of the variation within each sample, and to reduce it as much as possible.

We performed and analyzed a large dataset of 126 experiments in 96-well plate format involving 203 different fibroblast cell lines, out of which 26 were seeded in more than one plate (S1 Table). The large number of between-plate and within-plate replicates allowed us to statistically characterize the nature and magnitude of systematic and random variations in these data. We developed a statistical procedure called OCR-Stats, to extract robust and accurate oxygen consumption rates for each well, which translates into robust summarized values of the multiple replicates within and between plates. The OCR-Stats algorithm includes automatic outlier identification and controls for well and plate-interval effects, which led to a significant increase in accuracy over state-of-the-art methods.

Systematic and random variations were found to be multiplicative. This motivated us to establish bioenergetic measures based on differences in the logarithmic scale that translate into ratios and proportions in the natural scale: ETC-dependent OC proportion, ATPase-dependent OC proportion, ETC-dependent proportion of ATPase-independent OC, and maximal over initial OC fold change (Table 1).

Using an automatic outlier detection approach, we provide estimators for each instance and show empirically that they are normally distributed. This permitted the use of linear regression models for assessing the statistical significance of bioenergetic measure comparisons between two biological samples. Using positive and negative controls from individuals known to have mitochondrial respiratory defects, we show that OCR-Stats outperforms the currently used statistical tests, which suffer from an excess of either false positives (within-plate methods) or false negatives (between-plate methods).

Furthermore, our study provides experimental design guidance by i) showing that between-plate variation largely dominates within-plate variation, implying that it is important to seed the same cell lines in multiple plates, and ii) providing estimates of variances within and between plates for each bioenergetic measure allowing for statistical power computations. A free and open source implementation of OCR-Stats in the statistical language R is provided at [github.com/gagneurlab/OCR-Stats](https://github.com/gagneurlab/OCR-Stats).

## Results

### Experimental design and raw data

We measured the OCR, the ECAR, and the cell number of 203 dermal fibroblast cultures derived from patients suspected to suffer from rare mitochondrial diseases and control cells from healthy donors (normal human dermal fibroblasts: NHDF, Materials and methods, S1 Table). These were assayed in 126 plates, all using the same protocol (Materials and methods). Twenty-six cell lines were grown independently and were measured in multiple plates. We will refer to these growth replicates as different biological samples. The NHDF cell line was seeded in all the plates for the assessment of potential systematic plate effects. The corners of each plate were left as blank, that is, filled with media but not with cells, to control for changes in temperature [22]. One common layout of a plate is depicted in Fig 1C, showing how each biological sample is present in many well replicates. We seeded between 3 and 7 biological samples per plate (median = 4). This variation reflects typical set-ups of experiments in a lab performed over multiple years. Then, we used the standard mitochondrial stress test assay [20]

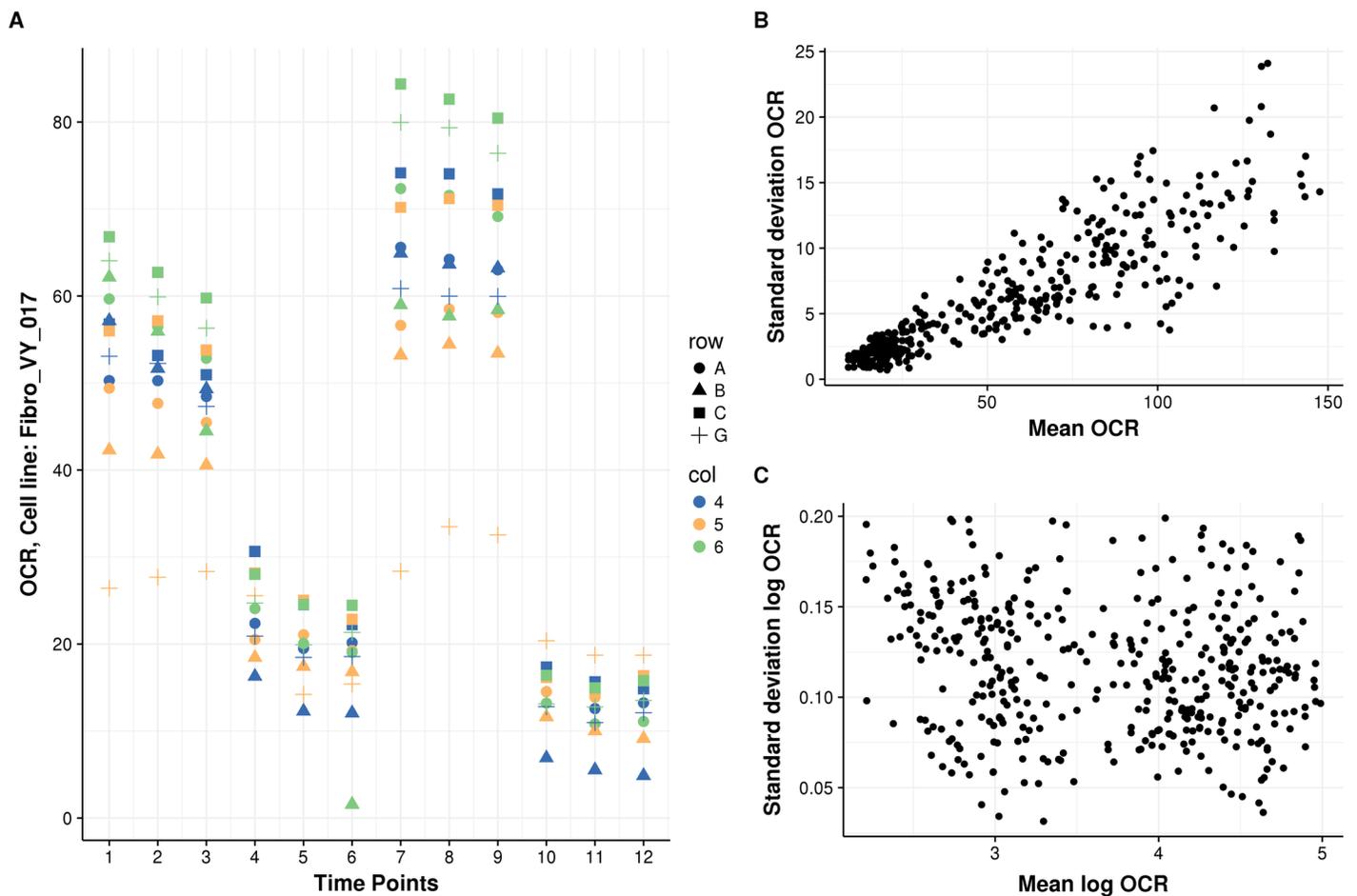
leading to four time intervals, with three time points each, denoted by  $Int_1$  (before adding any treatment),  $Int_2$  (after oligomycin),  $Int_3$  (after FCCP), and  $Int_4$  (after rotenone) (Fig 1A). In addition, we flagged wells that did not react as expected to the treatments and discarded them from the statistical analysis (Materials and methods).

### Variations between replicates within plates

Fig 2A shows representative replicate time series, with data from 12 wells for one biological sample in a single plate depicting commonly observed variations.

First, outlier data points occurred frequently. We distinguished two different types of outliers: entire series for a well (e.g., well G5 in Fig 2A) and individual data points (e.g., well B6 at time point 6 in Fig 2A). In the latter case, eliminating the entire series for well B6 would be too restrictive and would result in the loss of valuable data from the other 11 valid time points. Therefore, methods for detecting outliers taking these two possibilities into consideration must be devised.

Second, we noticed a proportional dependence of OCR value and standard deviation between replicates (Fig 2B), suggesting that the error is multiplicative. Unequal variance, or



**Fig 2. OCR behavior over time.** (A) Typical time series replicates inside a plate. Behavior of OCR expressed in pmol/min (y-axis) of Fibro\_VY\_017 over time (x-axis). Colors indicate the row and shape the column of 12-well replicates. Variation increases for larger OCR values, OCR has a systematic well effect, and there are two types of outliers: well-level and single-point. (B) Scatterplot of standard deviation (y-axis) vs. mean (x-axis) of all three time replicates of each interval, well, and plate of OCR of NHDF only shows a positive correlation ( $n = 409$ ). (C) The same as (B) but for the logarithm of OCR, where the correlation disappears.

<https://doi.org/10.1371/journal.pone.0199938.g002>

heteroscedasticity, can strongly affect the validity of statistical tests and the robustness of estimations. Therefore, we propose modeling OCR on a logarithmic scale, where the dependence between the variance and the mean disappears (Fig 2B and 2C). The activities of respiratory chain enzymes such as NADH-ubiquinone reductase also obey log-normal distributions [33].

Third, we observed systematic effects in OCR between wells (e.g., OCR values of well C6 are among the highest, while OCR values of well B5 are among the lowest at all the time points; Fig 2A). Variations in cell number, initial conditions, treatment concentrations, or fluorophore sleeve calibration can lead to systematic differences between wells, which we refer to as well effects. To investigate whether well effects could be corrected using cell number to a large extent as in [26], we counted the number of cells after the experiments using Cyquant (Materials and methods). As expected, the median OCR for each interval grows linearly with cell number measured at the end of the experiment (Spearman’s rho between 0.32 and 0.47,  $P < 2.2 \times 10^{-16}$ , S1A Fig). However, the relationship is not perfect, reflecting important additional sources of variations and also possible noise in measuring the cell number. Strikingly, dividing OCR by cell count led to a higher coefficient of variation (standard deviation divided by the mean) between the replicate wells than without that correction (S1B Fig). This analysis showed that normalization by the division of raw cell counts is insufficient and motivated us to derive another method to capture well effects. Finally, we found that sex does not significantly associate with OCR levels (S2 Fig), in agreement with [34].

### A statistical model for OCR within plates

Building on these insights, we introduced a statistical model for OCR within plates. For a given plate, we modeled the logarithm of OCR  $y_{w,t}$  of well  $w$  at time point  $t = 1, \dots, 12$  as a sum of time interval effects, well effects, and noise, that is:

$$y_{w,t} = \theta_{\text{biosample}(w),\text{interv}(t)} + \beta_w + \epsilon_{w,t}, \tag{1}$$

where  $\theta_{\text{biosample}(w),\text{interv}(t)}$  is the time interval effect of the biological sample in well  $w$  for interval  $\text{interv}(t) = 1, \dots, 4$  of time point  $t$  (Fig 1A),  $\beta_w$  is the relative effect of well  $w$  compared to the reference well, and  $\epsilon_{w,t}$  is the error.

We now present the OCR-Stats algorithm. For a given plate:

1. Fit the log linear model from Eq (1) using the least squares method, which consists of minimizing  $\sum_w \sum_t (y_{w,t} - \theta_{\text{biosample}(w),\text{interv}(t)} - \beta_w)^2$ , thus, obtaining the estimates  $\hat{\theta}_{\text{biosample}(w),\text{interv}(t)}$  (which correspond to:  $\theta_B, \theta_{A1}, \theta_{M1}, \theta_{E1}$ ; Fig 1A) and  $\hat{\beta}_w$ .
2. For each well  $w$  and time point  $t$  in interval  $i$ , compute the log OCR well deviations:  $d_{w,t} = y_{w,t} - \hat{\theta}_{\text{biosample}(w),\text{interv}(t)} - \frac{1}{n} \sum_w \hat{\beta}_w$ , which is used to identify both the well and the single point outliers (Materials and methods, S3 Fig).
3. Identify and remove well level outliers (Materials and methods). Fit again, iteratively, until no more are found (S3A and S3B Fig).
4. Identify and remove single point outliers (Materials and methods). Fit again, iteratively, until no more are found (S3C and S3D Fig).
5. Compute the ratio-based metrics (Table 1), or scale back to natural scale in order to compute the bioenergetic measures [e.g.: Basal respiration =  $\exp(\hat{\theta}_1) - \exp(\hat{\theta}_4)$ ].

Note that the well effect is modeled independently for each plate, that is, it corresponds to the effect of a well of a given plate and not to the effect of a well position shared across plates.

We investigated whether there were positional effects and found that OCR measurements are lower in the edges by a median of up to 13.1% (S4A and S4B Fig). However, these positional effects were consistent across intervals (S4A and S4B Fig). Consequently, these positional effects are to a large extent canceled (S4C and S4D Fig) when using the metrics that we suggest (Table 1) because they involve differences of log OCRs. One exception was row A, where median differences of up to 2.2% were observed for the ETC-dependent OCR proportion and for the maximal over initial OCR fold change (S4C and S4D Fig). Practitioners could avoid all four edges and not only the four corners as typically done. However, these systematic deviations are small compared to the amplitude of biological effects typically investigated (not less than 12%–30% [30–32]). Altogether, this approach led to coefficients of variation between wells of the same plate of 11%, 14%, 13%, and 17% for intervals 1, 2, 3 and 4 respectively (Materials and methods).

### Variations between plates

After analyzing the OCR variation among the wells inside plates, we aimed to study the variation across multiple plates. Using data from the controls NHDF, we found that the variability between plates in all the intervals is much larger than that between wells (S2 Table and S5 Fig). Variations between plates can arise, for example, due to differences in temperature, seeding time, growth time, growth medium, or sensor cartridge [13]. Moreover, treatment efficiencies can also vary between plates, but independently from each other. For example, the concentration of rotenone may differ in one plate. That would affect the OCR measurements of all the wells in that plate, but only in time interval 4.

Next, we investigated whether our assumption of systematic plate-interval effects held. We indeed observed a similar increase in OCR in interval 1 on both biological samples on plate #20140430 with respect to plate #20140428 (Fig 3A). To test whether this tendency held across the repeated biological samples, we compared all the replicate pairings with their respective NHDF controls and found a positive correlation in all the time intervals (Fig 3B), suggesting a plate-interval effect. These observations show the importance of basing conclusions from observations across multiple plates and for seeding a control cell line on every plate.

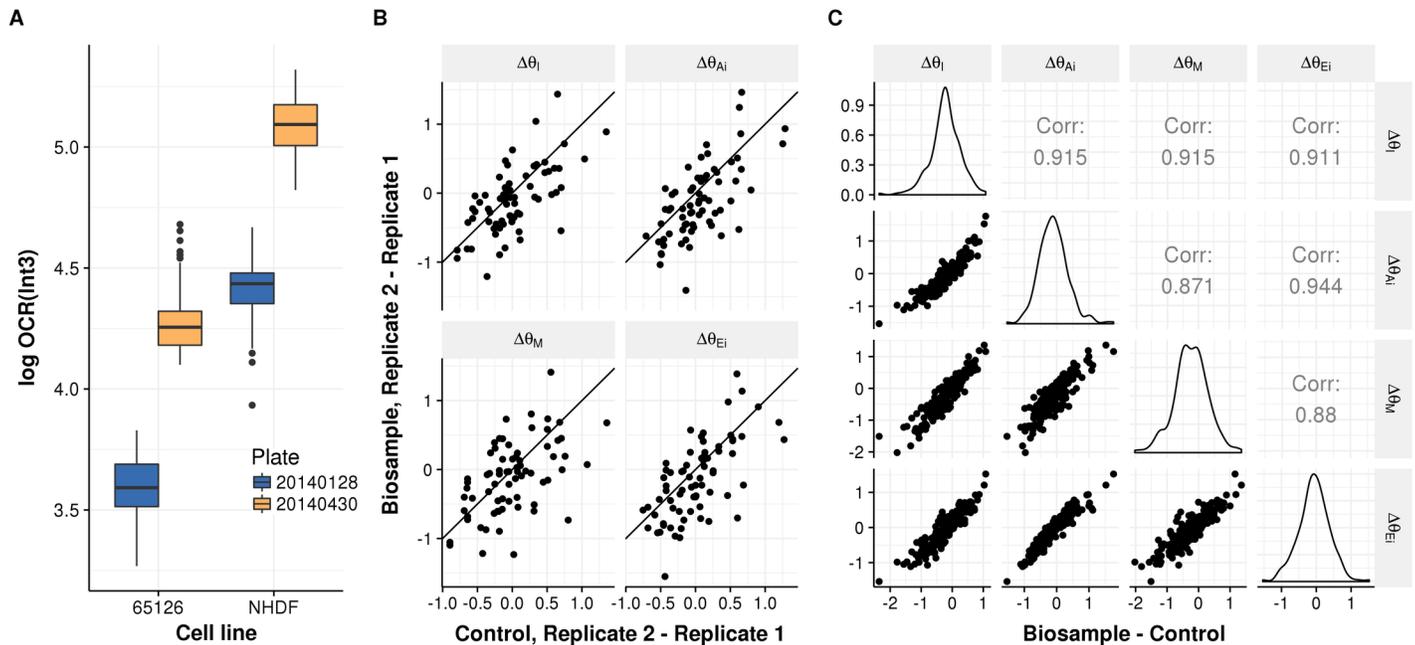
### Statistical testing for the comparison of biological samples across plates

We then set up to devise a model to statistically assess difference in OCR ratios between two biological samples across multiple plates. Since there is a remaining systematic effect across intervals at the plate level (Fig 3C) and because of the plate-interval effects, we recommend using ratios of OCR levels (i.e. differences in the logarithmic scale) (Table 2).

Subsequently, for any given OCR ratio (e.g., M/Ei-fold change, Tables 1 and 2), we test the differences of the OCR log-ratios  $\Delta\Delta\theta$  of a biological sample  $b$  versus a control  $c$  using the following linear model:

$$\Delta\Delta\theta_{b,p} = \mu_b + \epsilon_{b,p}, \tag{2}$$

where  $\Delta\Delta\theta_{b,p}$  is one OCR log-ratio difference of interest (Table 2) inside a plate  $p$ . We fit this model over our complete dataset using linear regression, thus obtaining one value  $\hat{\mu}_b$  per OCR ratio and biological sample  $b$ . Then, we tested these against the null hypothesis  $\mu_b = 0$  to compute p-values and confidence intervals (Materials and methods). Fitting the linear model of Equation (Eq 2) over the complete dataset gives a robust estimate of the standard deviation of the error term. Applying this approach, we found no evidence against the normality and homoscedasticity assumption of OCR-Stats as the quantile-quantile plots of the residuals aligned well along the diagonal (Fig 4A and S6 Fig).



**Fig 3. Plate-interval effect.** (A) Log of OCR in interval 3 (y-axis) for the cell lines #65126 and NHDF (x-axis), which were seeded in two different plates (color-coded). The similar increase in OCR from plate #20140128 to #20140430 in both biological samples suggests that there is a systematic plate-interval effect. (B) Scatterplots of the differences of the logarithm of OCR levels  $\Delta\theta$  of all possible 2 by 2 combinations of repeated biological samples across experiments (y-axis) against their respective controls (NHDF) (x-axis) showing that there is a positive correlation ( $I_1: \rho = 0.64, P < 2.3 \times 10^{-8}, I_2: \rho = 0.65, P < 1.2 \times 10^{-8}, I_3: \rho = 0.52, P < 1.2 \times 10^{-5}, I_4: \rho = 0.64, P < 1.4 \times 10^{-8}$ ), confirming a systematic plate-interval effect ( $n = 63$ ). (C) Scatterplot of the difference of log OCR levels  $\Delta\theta$  of all the biological samples vs. their respective control (both axes) of every interval with respect to another. All the differences  $\Delta\theta$  correlate with each other even after removing the plate-interval effect (by subtracting control values).

<https://doi.org/10.1371/journal.pone.0199938.g003>

### Benchmark

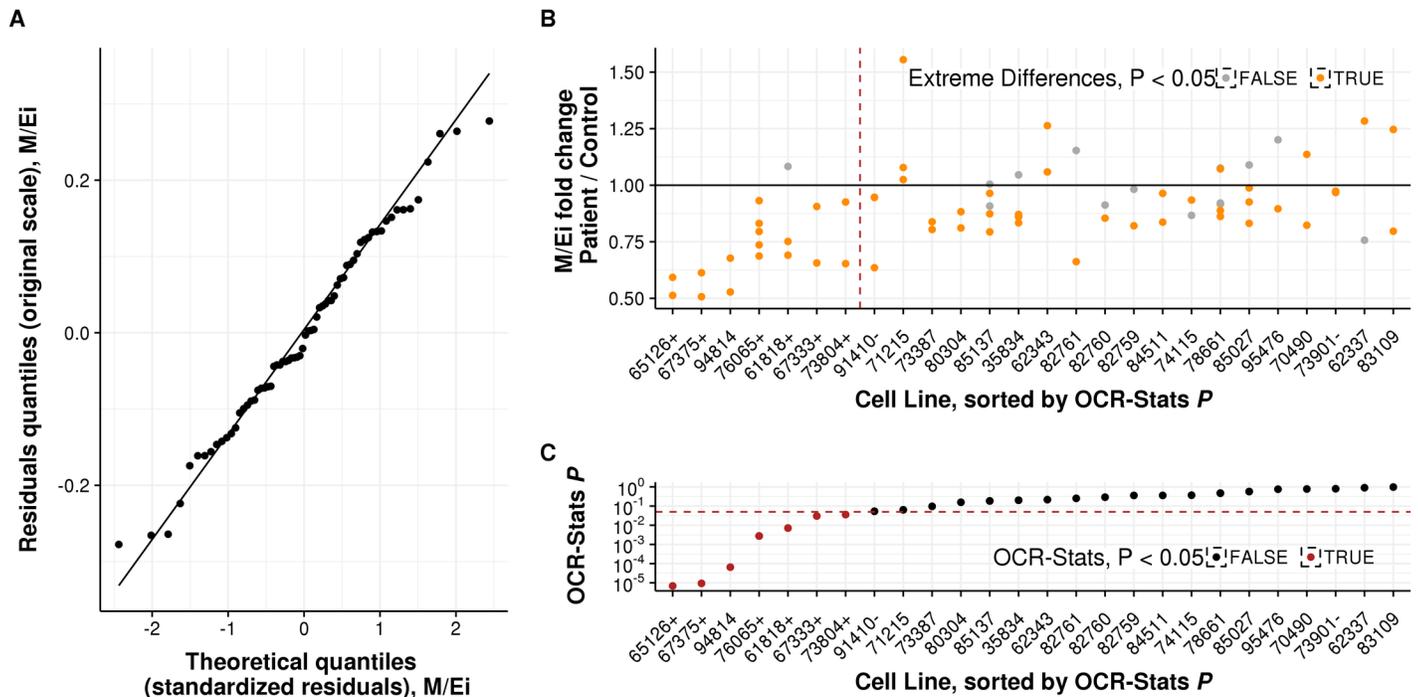
We applied OCR-Stats statistical testing, Extreme Differences plus Wilcoxon test within each plate (within-plate ED), and Extreme Differences plus Wilcoxon test across plates (across-plate ED) to obtain the M/Ei-fold change and maximal respiration (MR) of all of the 26 cell lines that were seeded in more than one plate (Materials and methods). For every approach, we computed p-values for significant fold changes against the controls. Six of these cell lines are derived from patients with rare variants in genes associated with an established cellular respiratory defect, allowing the assessment of the sensitivity (or statistical power) of each approach (S3 Table, [35–39]). Additionally, two cell lines (#73901 and #91410) repeatedly showed no significant respiratory defects in earlier studies and served as negative controls [40,41].

**Table 2. OCR ratio-based differences for statistical testing.**

OCR ratios	Tested differences $\Delta\Delta\theta$
E/I-proportion	$(\theta_{I,b} - \theta_{Ei,b}) - (\theta_{I,c} - \theta_{Ei,c})$
A/I-proportion	$(\theta_{I,b} - \theta_{Ai,b}) - (\theta_{I,c} - \theta_{Ai,c})$
E/Ai-proportion	$(\theta_{Ai,b} - \theta_{Ei,b}) - (\theta_{Ai,c} - \theta_{Ei,c})$
M/I-fold change	$(\theta_{M,b} - \theta_{I,b}) - (\theta_{M,c} - \theta_{I,c})$
M/Ei-fold change	$(\theta_{M,b} - \theta_{Ei,b}) - (\theta_{M,c} - \theta_{Ei,c})$

For each OCR ratio from Table 1, we present the differences  $\Delta\Delta\theta$  to be used when testing a biological sample  $b$  against a control  $c$  on each plate.

<https://doi.org/10.1371/journal.pone.0199938.t002>

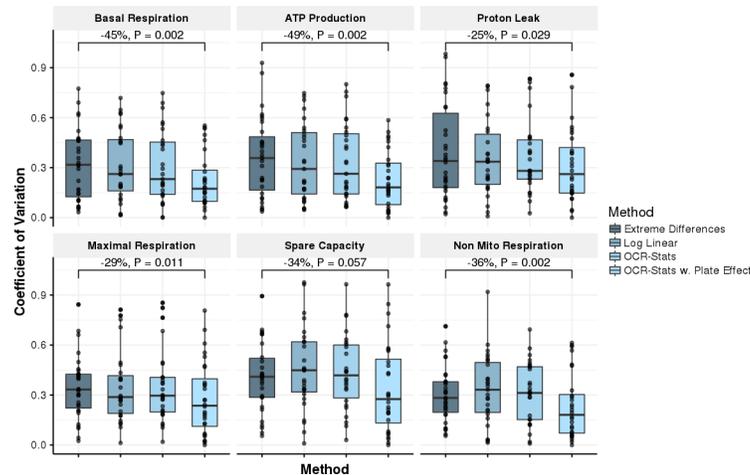


**Fig 4. Statistical testing of M/Ei-fold change patient vs. control on multiple plates.** (A) Ratio of M/Ei-fold change (y-axis) of all the cell lines repeated across plates (x-axis) and their respective controls, sorted by the p-value obtained using the OCR-Stats method. Left of the red dashed line are cell lines with significantly lower M/Ei-fold change using OCR-Stats. Dots in orange represent biological samples with significantly lower or higher M/Ei-fold change using the ED method. Highlighted positive (+) and negative (-) controls. (B) Similar to (A), but depicting the p-value in logarithmic scale (y-axis) using OCR-Stats. Red dashed line at  $P = 0.05$ . Dots in red represent cell lines with significantly lower M/Ei-fold change using the OCR-Stats method. (C) Quantile-quantile theoretical (x-axis) vs. observed (y-axis) plot of the residuals  $\epsilon_{b,p}$  of the linear model (2) applied to M/Ei-fold change. Points are lying on the diagonal as expected from normally distributed residuals.

<https://doi.org/10.1371/journal.pone.0199938.g004>

The within-plate ED method reported significantly higher or lower MR for 56 out of 69 (81.2%) biological samples with respect to the control (Fig 4B and S3 Table). Moreover, the within-plate ED method reported one or more significant differences for all the 26 cell lines, and one or more non-significant differences for 11 cell lines (Fig 4B). For two cell lines, the within-plate ED method returned significant differences with opposite signs (cell lines #78661, #83109, Fig 4B). These ambiguous results show the importance of testing using multiple plates and suggest the need for a more robust approach than the within-plate ED. One approach to evaluate samples measured in multiple plates is to perform a Wilcoxon test on the ED values averaged per plate (across-plate ED, Materials and methods). However, this requires at least five plate replicates in order to obtain significant results. Here, one cell line only, #78661, was found to have significantly impaired OCR in this way. For these data, OCR-Stats was much more conservative than within-plate ED and found only 7 out of 26 (26.9%) cell lines to have aggregated significantly lower M/Ei-fold change than the control, including all six positive control cell lines (Fig 4B and 4C, and S3 Table). Moreover, OCR-Stats did not report significant M/Ei-fold changes for the two negative controls.

Furthermore, we computed the coefficient of variation (standard deviation divided by mean) of the six bioenergetic measures in the natural scale (Table 1) of all the repeated biological samples across plates for the following methods: i) the default Extreme Differences (ED) method (Materials and methods) provided by the vendor, ii) the log linear (LL) corresponding to steps 1 and 2 of the OCR-Stats algorithm, iii) complete OCR-Stats (LL + outlier removal), and iv) OCR-Stats after correcting for plate effect (OCR-PE) using Eq (4) (Materials and



**Fig 5. Benchmark using coefficient of variation.** Coefficient of variation (CV = standard deviation/mean, y-axis) of replicates across experiments ( $n = 26$ ) using different methods (x-axis) to estimate the six bioenergetic measures. In all, except for Spare Capacity, OCR-Stats with plate-interval effect showed significantly lower variation with respect to the Extreme Differences method. P-values obtained from one-sided paired Wilcoxon test.

<https://doi.org/10.1371/journal.pone.0199938.g005>

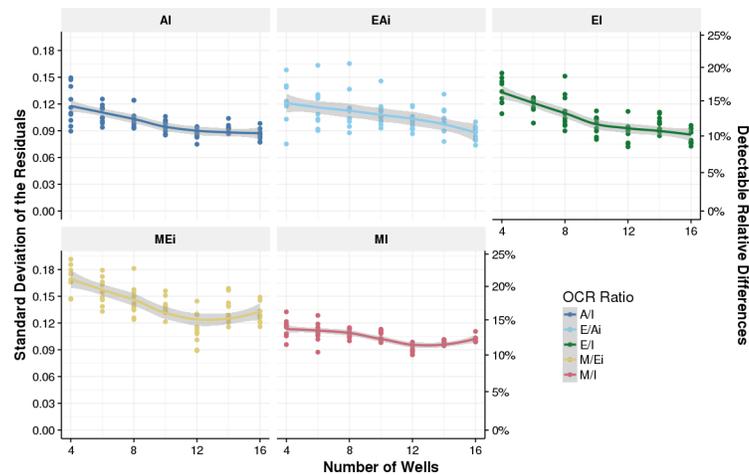
methods). Each step contributed to a decrease in the coefficient of variation, obtaining a final significant reductions of 45% and 29% in basal and maximal respiration, respectively, from plate-corrected OCR-Stats (OCR-PE) with respect to ED ( $P < 0.012$ , one-sided Wilcoxon test) (Fig 5). Taken together, these results show that OCR-Stats successfully identifies and decreases the variation within and between plates, providing more stable testing results, which translates into fewer false positives.

### Power analysis

Finally, we investigated the statistical power of OCR-Stats in this dataset. Specifically, we are interested in determining the minimum relative differences our method is able to significantly detect, and the minimal number of well replicates needed. We subsetted the number of wells of the repeated biological samples to 4, 6, 8, 10, 12, 14, and 16 wells on each plate, and used the OCR-Stats algorithm (including outlier removal) and statistical testing to obtain the residuals  $\epsilon_{b,p}$  and their standard deviation (Fig 6). Assuming three plates per comparison and 16 wells per plate, these standard deviations allow detecting relative differences of 10% to 15% depending on the considered log OCR ratios differences for significance level of 5% (Fig 6, right y-axis, Materials and methods). Relative differences of 10% to 15% are in line with reported detected variations in the literature which we found to be as low as 12%-30% [30–32]. This analysis also suggests to seed at least 12 wells per biological sample per plate, since we observed increased standard deviations of the residuals for numbers of wells smaller than 12. Note that this power calculation is based on measurements performed in our laboratory only. Other laboratories might have larger or smaller measurement variations. Nonetheless, our procedure could be used as a guideline for power calculation.

### Discussion and conclusion

Mitochondrial studies using extracellular fluxes, specifically the XF Analyzer from Seahorse, are gaining popularity and are finding their way into diagnostics; therefore, it is of paramount importance to have an appropriate statistical method to estimate the OCR levels from the raw



**Fig 6. Power analysis.** Standard deviation of the residuals from the model in Eq (2) (left y-axis) against number of wells per biological sample and per plate (x-axis) for each OCR log-ratio difference (Table 1). The right y-axis corresponds to the minimal detectable relative differences using three plates at 5% significance level (Materials and methods). For every number of wells, the 10 data points correspond to each of the 10 random samplings without replacement of the wells per biological sample and per plate.

<https://doi.org/10.1371/journal.pone.0199938.g006>

data. Here, we have developed such a model, the OCR-Stats algorithm, which includes approaches to control for well and plate-interval effects, and automatic outlier identification.

We found that dividing cellular OCR by cell number involved the introduction of more noise than was seen for uncorrected data. Here, we always seeded the same number of cells. Hence, the variations across wells that we observed in the cell number at the end of the experiments are largely overestimated by noise in the measurements. In other experimental settings in which different numbers of cells are seeded, we suggest the inclusion of an offset term to the model in Eq (1) equal to the logarithm of the seeded cell number to control for this variation by design. In addition, the Seahorse XF Analyzer can be used on isolated mitochondria and on isolated enzymes, where a normalization approach is to divide OCR by mitochondrial proteins or enzyme concentration [42]. However, as described here for cellular assays, robust normalization procedures require careful analysis.

We demonstrated that OCR comparisons should be performed using ratios rather than using differences, and that the cell lines must be seeded on the same plate, as this eliminates sources of variation like cell number, and well positional and plate-interval effects. We introduced a linear model, the OCR-Stats statistical testing, and showed that the results agree with previous results of patients diagnosed with mitochondrial disorders. We showed that the variation in differences of OCR log-ratios  $\Delta\Delta\theta_{b,p}$  for the same biological sample across plates is large, and that, consequently, samples should be seeded in multiple plates. Note that a contaminated sample can increase the variability, affecting the significance of all the other samples. Therefore, it is important to detect such samples and to exclude them from further analysis. By doing power analysis, we showed that our method is able to detect relative differences of 10% - 15%, and that the minimum number of well replicates per biological sample in a 96-well plate should be 12.

We encourage users to consider all five metrics (Table 1). Severely affected cell lines with strongly reduced E/I-proportion might not necessarily show a clear effect on M/Ei-fold change. For example, cell line #73387 was found to have a lower, but not significantly ( $P < 0.10$ ), M/Ei-fold change (the most common metric used throughout the literature, Fig 4C and S3 Table), but when analyzing its E/I proportion, we found that it was significantly lower

than the control ( $P < 1.2 \times 10^{-7}$ ). This result is consistent with its genetic diagnosis, a homozygous loss of function variant in the *PET100* gene, which is involved in biogenesis of mitochondrial complex IV (S1 Table, [43]).

In principle, OCR-Stats should be able to estimate ECAR levels. To guarantee that the method is indeed applicable, similar analyses as performed here should be done beforehand. Preliminary investigations suggest that the nature of noise (outliers, multiplicative) is similar to that for OCR.

## Materials and methods

### Biological material

All the biological samples were derived from primary fibroblast cell lines of humans suffering from rare mitochondrial diseases, established in the framework of the German and European networks for mitochondrial disorders mitoNet and GENOMIT. All the individuals or their guardians provided written informed consent for their cell lines to be used for evaluation and testing, in agreement with the Declaration of Helsinki and approved by the ethical committees of the centers participating in this study. All the assays were performed in accordance with the local approval of the ethical committee of the Technical University of Munich. The controls are primary patient fibroblast cell lines, normal human dermal fibroblasts (NHDF) from neonatal tissue, commercially available from Lonza, Basel, Switzerland.

### Measure of extracellular fluxes using Seahorse XF96

We seeded 20,000 fibroblast cells in each well of a XF 96-well cell culture microplate in 80 ml of culture medium, and incubated them overnight at 37°C in 5% CO<sub>2</sub>. The four corners were left only with medium for background correction. The culture medium was replaced with 180 ml of bicarbonate-free DMEM and cells were incubated at 37°C for 30 min before measurement. Oxygen consumption rates (OCR) were measured using an XF96 Extracellular Flux Analyzer [20]. OCR was determined at four levels: with no additions, and after adding oligomycin (1 μM), carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP, 0.4 μM), and rotenone (2 μM) (additives purchased from Sigma at highest available quality). After each assay, manual inspection was performed on all wells using a conventional light microscope. The wells for which the median OCR level did not follow the expected order, namely, median[OCR(Int<sub>3</sub>)] > median[OCR(Int<sub>1</sub>)] > median[OCR(Int<sub>2</sub>)] > median[OCR(Int<sub>4</sub>)] (Fig 1A), were discarded (977 wells, 10.47%). It is important to note that other cell lines, or cell lines under certain conditions, may not react as expected to the standard treatments; therefore, they should not be discarded. In addition, we excluded contaminated wells and wells in which the cells got detached (461 wells, 4.94%) from the analysis. All the raw OCR data are available in S4 Table.

### Cell number quantification

The cell number was quantified using the CyQuant Cell Proliferation Kit (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's protocol. In brief, the cells were washed with 200 μL PBS per well and frozen in the microplate at -80°C to ensure subsequent cell lysis. The cells were thawed and resuspended vigorously in 200 μL of 1x cell-lysis buffer supplemented with 1x CyQUANT GR dye per well. The resuspended cells were incubated in the dark for 5 min at RT, whereupon fluorescence was measured (excitation: 480 nm, emission: 520 nm).

### Extreme differences (default) method to compute bioenergetic measures

On every plate independently, for each well, in interval 1 take the OCR corresponding to the last measurement, in intervals 2 and 4 take the minimum, and in interval 3 the maximum OCR value [19]. Then, use the corresponding differences to estimate the bioenergetic measures. Report the results per patient as the mean across wells plus standard deviation or standard error, separately for each plate.

### Outlier removal

For each sample  $s$  and well  $w$ , compute the mean across time points of its squared deviations:  $s_w := \text{mean}_t(d_{w,t}^2)$ , thus, obtaining a vector  $\mathbf{s}$ . Identify as outliers the wells whose  $s_w > \text{median}(\mathbf{s}) + 5 \text{mad}(\mathbf{s})$ , where  $\text{mad}$ , median absolute deviation, is a robust estimation of the standard deviation (S3A Fig). We found that deviations by 5  $\text{mad}$  from the median were sufficiently selective in practice. Compute the vector of estimates  $\hat{\theta}$  using the remaining wells and iterate this procedure until no more wells are identified as outliers. It required eight iterations until convergence and around 16.5% of all the wells were found to be outliers (S3B Fig).

Single point outliers are identified in a similar way. After discarding the wells that were found to be outliers in the previous step, categorize as outliers single data points whose  $d_{w,t}^2 > \text{median}_t(d_{w,t}^2) + 7 \text{mad}_t(d_{w,t}^2)$  (S3C Fig). Iterate until no more outliers are found. It required 19 iterations until convergence and approximately 6.1% of single points were found to be outliers (S3D Fig).

### Coefficient of variation between wells of the same plate

Using only the controls NHDF, we computed the standard deviation  $\sigma_{p,i}$  of the logarithm of OCR across all the wells for each plate  $p$  and interval  $i$ . Then, we computed the median across plates, thus, obtaining one value  $\bar{\sigma}_i$  per interval ( $\bar{\sigma}_1 = 0.10, \bar{\sigma}_2 = 0.13, \bar{\sigma}_3 = 0.12, \bar{\sigma}_4 = 0.16$ ). Coefficients of variation in the natural scale were approximated by taking the exponential of these standard deviations.

### OCR-Stats statistical testing

We fitted Eq (2) using linear regression as implemented in the base R function `lm()`. P-values for each ratio against the null hypothesis  $\mu_b = 0$  are obtained with the default test (Student's t-test) returned by the summary function on the `lm` fit object.

### Power calculation of multi-plate experiments

Minimal detectable effects for OCR-ratio based metrics (Table 2) at 95% confidence level were estimated using the following equation:

$$\exp\left(1.96 \frac{\text{sd}(\epsilon_{b,p})}{\sqrt{n}}\right) - 1, \tag{3}$$

where  $\epsilon_{b,p}$  are the residuals from Eq 3 and 1.96 corresponds to the approximate value of the 97.5 percentile of the standard normal distribution. We obtain the metrics on Table 2 by setting  $n = 3$ .

### Plate-interval effect benchmark

For benchmarking, we correct for the plate-interval effect using only the data from the controls NHDF  $c$  of each plate using the following model:

$$y_{control,t,p} = \theta_{control,interv(t)} + \beta_{interv(t),p} + \varepsilon_{t,p}. \quad (4)$$

We solved it using the least squares method and used the effects  $\hat{\beta}_{i,p}$  as offsets in Eq (1). We recomputed  $\hat{\theta}_{b,i}$  values accordingly and scaled back to the natural scale to calculate the bioenergetic measures and the coefficient of variation of all repeated the biological samples, except the control (Fig 5).

### Multi-plate averaging method

In the case of inter-plate comparisons, the multi-plate averaging method takes the mean and standard error of the bioenergetic measures obtained using the Extreme Differences (ED) method of all the repeated biological samples across plates [44].

### Supporting information

**S1 Table. Sample metadata.** Each row corresponds to a different fibroblast cell line (Fibroblast id) from a patient with a mitochondrial disorder. In cases in which the causal pathogenic gene was found and validated, it appears in the Gene column with its respective OMIM number. N indicates the number of experiments of each sample.

(TXT)

**S2 Table. Coefficient of variation within and between plates.** Coefficient of variation computed as mean/standard deviation of OCR within and between plates using the controls NHDF only, in each time interval.

(TXT)

**S3 Table. M/Ei-fold change and maximal respiration (MR) differences of samples repeated across experiments with respective p-values.** Each row corresponds to a different biological sample (cell\_culture), with the difference in M/Ei-fold change and MR with respect to the control NHDF, in single and multiple plates. P-values computed using OCR-Stats and ED method within and between plates.

(TXT)

**S4 Table. OCR raw data from all experiments.** OCR and cell number raw data for each of the 126 plates and 203 samples across all the 12 time points and 4 treatments.

(TXT)

**S1 Fig. Normalizing by cell number does not reduce variation.** (A) OCR per well median (y-axis) vs. cell number (in thousands, x-axis) of the controls NHDF in all experiments ( $n = 2,192$  for each panel) showing that there is a positive correlation in all the time intervals ( $I_1: \rho = 0.47$ ,  $I_2: \rho = 0.45$ ,  $I_3: \rho = 0.40$ ,  $I_4: \rho = 0.33$ ;  $P < 2.2 \times 10^{-16}$  for all the intervals). (B) Coefficient of variation (y-axis) of well replicates within plates for raw OCR and OCR normalized dividing by cell count (x-axis), split for each time interval. Each point represents a different sample. In all the four intervals, not only did normalization not reduce the coefficient of variation, but it actually increased it. P-values obtained from two-tailed Wilcoxon tests.

(PNG)

**S2 Fig. OCR does not depend on sex.** OCR levels  $\theta$  (y-axis) split by sex (x-axis). We see no significant difference in any time interval ( $n = 45$  male, 30 female).

(PNG)

**S3 Fig. Outlier detection.** (A) Number of wells (y-axis) identified as outliers on each iteration (x-axis). Around 16.5% of all valid wells detected as outliers. (B) Mean (per well) squared errors distribution for cell line Fibro\_VY\_014. Wells beyond the red line (median +  $5 \times \text{mad}$ ) are recognized as well-level outliers. (C) Number of single-point outliers (y-axis) identified on each iteration (x-axis). Around 6.1% of remaining data (after removing well outliers) detected as single point outliers. (D) Squared error distribution for cell line Fibro\_VY\_076. Points beyond the red dashed line (median +  $7 \times \text{mad}$ ) are recognized as single-point outliers.

(PNG)

**S4 Fig. Investigation of location effect.** (A) Deviations of the log OCR measurements with respect to the interval effect ( $y - \hat{\theta}$ , y-axis) behavior across rows (x-axis). In general, a tendency for higher OCR is observed on the center of the plate across all time intervals. (B) The same as (A) but for columns (x-axis). Lower values observed in the edges. (C, D) Well-level OCR ratio subtracted interval level OCR ratio (Table 1) across rows (x-axis, C) and columns (x-axis, D).

All the location effects get canceled, except for row A where it remains relatively low.

(PNG)

**S5 Fig. Variation between plates is larger than variation within plates.** Boxplot showing OCR in time interval 3 (x-axis) of NHDF seeded in 10 randomly selected plates (y-axis) reflecting that the variation between is larger than the variation within plates. Red line: mean of OCR across all plates. This trend was observed across all the plates and for all the intervals (S2 Table).

(PNG)

**S6 Fig. Residuals from the linear regression are consistent with a normal distribution.**

Quantile-quantile theoretical (x-axis) vs. observed (y-axis) plots of the residuals  $\epsilon_{b,p}$  of the linear model from Eq (2). Points lie on the diagonal as expected from normally distributed residuals.

(PNG)

## Acknowledgments

We would like to thank Daniel Bader, Žiga Avsec, Jun Cheng, and all the members of the Gagneur Lab for the valuable discussions and manuscript revision.

## Author Contributions

**Conceptualization:** Vicente A. Yépez, Laura S. Kremer, Arcangela Iuso, Mirjana Gusic, Robert Kopajtich, Eliška Koňářiková, Agnieszka Nadel, Holger Prokisch, Julien Gagneur.

**Data curation:** Laura S. Kremer, Arcangela Iuso, Mirjana Gusic, Robert Kopajtich, Eliška Koňářiková, Agnieszka Nadel.

**Formal analysis:** Vicente A. Yépez, Holger Prokisch, Julien Gagneur.

**Investigation:** Vicente A. Yépez, Julien Gagneur.

**Software:** Vicente A. Yépez, Leonhard Wachutka.

**Supervision:** Holger Prokisch, Julien Gagneur.

**Visualization:** Vicente A. Yépez, Julien Gagneur.

**Writing – original draft:** Vicente A. Yépez, Holger Prokisch, Julien Gagneur.

**Writing – review & editing:** Vicente A. Yépez, Laura S. Kremer, Arcangela Iuso, Mirjana Gusic, Robert Kopajtich, Eliška Koňářková, Agnieszka Nadel, Leonhard Wachutka, Holger Prokisch, Julien Gagneur.

## References

1. Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, et al. Mitochondrial diseases. *Nat Rev Dis Prim*. Macmillan Publishers Limited; 2016; 2. <https://doi.org/10.1038/nrdp.2016.80> PMID: [27775730](https://pubmed.ncbi.nlm.nih.gov/27775730/)
2. Bholá PD, Letai A. Mitochondria—Judges and Executioners of Cell Death Sentences. *Mol Cell*. Elsevier Inc.; 2016; 61: 695–704. <https://doi.org/10.1016/j.molcel.2016.02.019> PMID: [26942674](https://pubmed.ncbi.nlm.nih.gov/26942674/)
3. Sun N, Youle RJ, Finkel T. The Mitochondrial Basis of Aging. *Mol Cell*. Elsevier Inc.; 2016; 61: 654–666. <https://doi.org/10.1016/j.molcel.2016.01.028> PMID: [26942670](https://pubmed.ncbi.nlm.nih.gov/26942670/)
4. Wallace DC. Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annu Rev Biochem*. 2007; 76: 781–821. <https://doi.org/10.1146/annurev.biochem.76.081205.150955> PMID: [17506638](https://pubmed.ncbi.nlm.nih.gov/17506638/)
5. Birsoy K, Wang T, Chen WW, Freinkman E, Abu-Remaileh M, Sabatini DM. An Essential Role of the Mitochondrial Electron Transport Chain in Cell Proliferation Is to Enable Aspartate Synthesis. *Cell*. 2015; 162: 540–551. <https://doi.org/10.1016/j.cell.2015.07.016> PMID: [26232224](https://pubmed.ncbi.nlm.nih.gov/26232224/)
6. Sullivan LB, Gui DY, Hosios AM, Bush LN, Freinkman E, Vander Heiden MG. Supporting Aspartate Biosynthesis Is an Essential Function of Respiration in Proliferating Cells. *Cell*. Elsevier Inc.; 2015; 162: 552–563. <https://doi.org/10.1016/j.cell.2015.07.017> PMID: [26232225](https://pubmed.ncbi.nlm.nih.gov/26232225/)
7. Zong W-X, Rabinowitz JD, White E. Mitochondria and Cancer. *Mol Cell*. Elsevier Inc.; 2016; 166: 555–566. <https://doi.org/10.1016/j.cell.2016.07.002> PMID: [27471965](https://pubmed.ncbi.nlm.nih.gov/27471965/)
8. Weinberg SE, Sena LA, Chandel NS. Mitochondria in the regulation of innate and adaptive immunity. *Immunity*. Elsevier Inc.; 2015; 42: 406–417. <https://doi.org/10.1016/j.immuni.2015.02.002> PMID: [25786173](https://pubmed.ncbi.nlm.nih.gov/25786173/)
9. Titov D V, Cracan V, Goodman RP, Peng J, Grabarek Z, Mootha VK. Complementation of mitochondrial electron transport chain by manipulation of the NAD<sup>+</sup>/NADH ratio. *Science (80-)*. 2016; 352: 231–235. <https://doi.org/10.1126/science.aad4017> PMID: [27124460](https://pubmed.ncbi.nlm.nih.gov/27124460/)
10. Wallace DC. Mitochondria and cancer. *Nat Rev Cancer*. Nature Publishing Group; 2012; 12: 685–698. <https://doi.org/10.1038/nrc3365> PMID: [23001348](https://pubmed.ncbi.nlm.nih.gov/23001348/)
11. Dunham-Snary KJ, Sandel MW, Westbrook DG, Ballinger SW. Redox Biology A method for assessing mitochondrial bioenergetics in whole white adipose tissues. *Redox Biol*. Elsevier Ltd.; 2014; 2: 656–660. <https://doi.org/10.1016/j.redox.2014.04.005> PMID: [24936439](https://pubmed.ncbi.nlm.nih.gov/24936439/)
12. Yao J, Irwin RW, Zhao L, Nilsen J, Hamilton RT, Brinton RD. Mitochondrial bioenergetic deficit precedes Alzheimer's pathology in female mouse model of Alzheimer's disease. *PNAS*. 2009; 106: 14670–14675. <https://doi.org/10.1073/pnas.0903563106> PMID: [19667196](https://pubmed.ncbi.nlm.nih.gov/19667196/)
13. Koopman M, Michels H, Dancy BM, Kamble R, Mouchiroud L, Auwerx J, et al. A screening-based platform for the assessment of cellular respiration in *Caenorhabditis elegans*. *Nat Protoc*. Nature Publishing Group; 2016; 11: 1798–1816. <https://doi.org/10.1038/nprot.2016.106> PMID: [27583642](https://pubmed.ncbi.nlm.nih.gov/27583642/)
14. Dmitriev RI, Papkovsky DB. Optical probes and techniques for O<sub>2</sub> measurement in live cells and tissue. *Cell Mol Life Sci*. 2012; 69: 2025–2039. <https://doi.org/10.1007/s00018-011-0914-0> PMID: [22249195](https://pubmed.ncbi.nlm.nih.gov/22249195/)
15. Brand MD, Nicholls DG. Assessing mitochondrial dysfunction in cells. *Biochem J*. 2011; 435: 297–312. <https://doi.org/10.1042/BJ20110162> PMID: [21726199](https://pubmed.ncbi.nlm.nih.gov/21726199/)
16. Hill BG, Benavides GA, Jr JRL, Ballinger S, Italia LD. Integration of cellular bioenergetics with mitochondrial quality control and autophagy. *Biol Chem*. 2012; 393: 1485–1512. <https://doi.org/10.1515/hsz-2012-0198> PMID: [23092819](https://pubmed.ncbi.nlm.nih.gov/23092819/)
17. Gerencser AA, Neilson A, Choi SW, Edman U, Yadava N, Oh RJ, et al. Quantitative Microplate-Based Respirometry with Correction for Oxygen Diffusion. *Anal Chem*. 2009; 81: 6868–6878. <https://doi.org/10.1021/ac900881z> PMID: [19555051](https://pubmed.ncbi.nlm.nih.gov/19555051/)
18. Ribeiro SM, Giménez-cassina A, Danial NN. Measurement of Mitochondrial Oxygen Consumption Rates in Mouse Primary Neurons and Astrocytes. *Methods Mol Biol*. 2015; 1241: 59–69. [https://doi.org/10.1007/978-1-4939-1875-1\\_6](https://doi.org/10.1007/978-1-4939-1875-1_6) PMID: [25308488](https://pubmed.ncbi.nlm.nih.gov/25308488/)

19. Divakaruni AS, Paradyse A, Ferrick DA, Murphy AN, Jastroch M. Analysis and interpretation of microplate-based oxygen consumption and pH data. *Methods in Enzymology*. 2014. <https://doi.org/10.1016/B978-0-12-801415-8.00016-3>
20. Agilent Technologies. Mito Stress Test Kit, User Guide. 2017; Available: [https://www.agilent.com/cs/library/usermanuals/public/XF\\_Cell\\_Mito\\_Stress\\_Test\\_Kit\\_User\\_Guide.pdf](https://www.agilent.com/cs/library/usermanuals/public/XF_Cell_Mito_Stress_Test_Kit_User_Guide.pdf)
21. Ferrick DA, Neilson A, Beeson C. Advances in measuring cellular bioenergetics using extracellular flux. *Drug Discov Today*. 2008; 13. <https://doi.org/10.1016/j.drudis.2007.12.008> PMID: 18342804
22. Dranka BP, Benavides GA, Diers AR, Giordano S, Blake R, Reilly C, et al. Assessing bioenergetic function in response to oxidative stress by metabolic profiling. *Free Radic Biol Med*. 2011; 51: 1621–1635. <https://doi.org/10.1016/j.freeradbiomed.2011.08.005> PMID: 21872656
23. Zhang J, Nuebel E, Wisidagama DRR, Setoguchi K, Hong JS. Measuring energy metabolism in cultured cells, including human pluripotent stem cells and differentiated cells. *Nat Protoc*. 2012;7. <https://doi.org/10.1038/nprot.2012.048>.Measuring
24. Zhou W, Choi M, Margineantu D, Margaretha L, Hesson J, Cavanaugh C, et al. HIF1 $\alpha$  induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J*. Nature Publishing Group; 2012; 31: 2103–2116. <https://doi.org/10.1038/emboj.2012.71> PMID: 22446391
25. Shah-Simpson S, Pereira CFA, Dumoulin PC, Caradonna KL, Burleigh BA. Bioenergetic profiling of *Trypanosoma cruzi* life stages using Seahorse extracellular flux technology. *Mol Biochem Parasitol*. Elsevier B.V.; 2016; 208: 91–95. <https://doi.org/10.1016/j.molbiopara.2016.07.001> PMID: 27392747
26. Dranka BP, Hill BG, Darley-Usmar VM. Mitochondrial reserve capacity in endothelial cells: The impact of nitric oxide and reactive oxygen species. *Free Radic Biol Med*. Elsevier Inc.; 2010; 48: 905–914. <https://doi.org/10.1016/j.freeradbiomed.2010.01.015> PMID: 20093177
27. Chacko BK, Kramer P a, Ravi S, Benavides G a, Mitchell T, Dranka BP, et al. The Bioenergetic Health Index: a new concept in mitochondrial translational research. *Clin Sci*. 2014; 127: 367–373. <https://doi.org/10.1042/CS20140101> PMID: 24895057
28. Invernizzi F, D'Amato I, Jensen PB, Ravaglia S, Zeviani M, Tiranti V. Microscale oxygraphy reveals OXPHOS impairment in MRC mutant cells. *Mitochondrion*. Elsevier B.V. and Mitochondria Research Society. All rights reserved; 2012; 12: 328–335. <https://doi.org/10.1016/j.mito.2012.01.001> PMID: 22310368
29. Zhang J, Khvorostov I, Hong JS, Oktay Y, Vergnes L, Nuebel E, et al. UCP2 regulates energy metabolism and differentiation potential of human pluripotent stem cells. *EMBO J*. Nature Publishing Group; 2011; 30: 4860–4873. <https://doi.org/10.1038/emboj.2011.401> PMID: 22085932
30. Stroud DA, Surgenor EE, Formosa LE, Reljic B, Frazier AE, Dibley MG, et al. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature*. Nature Publishing Group; 2016; 538: 1–17. <https://doi.org/10.1038/nature19754> PMID: 27626371
31. Mitsopoulos P, Chang Y-H, Wai T, König T, Dunn SD, Langer T, et al. Stomatin-like protein 2 is required for in vivo mitochondrial respiratory chain supercomplex formation and optimal cell function. *Mol Cell Biol*. 2015; 35: 1838–47. <https://doi.org/10.1128/MCB.00047-15> PMID: 25776552
32. Almontashiri NAM, Chen HH, Mailloux RJ, Tatsuta T, Teng ACT, Mahmoud AB, et al. SPG7 Variant Escapes Phosphorylation-Regulated Processing by AFG3L2, Elevates Mitochondrial ROS, and Is Associated with Multiple Clinical Phenotypes. *Cell Rep*. 2014; 7: 834–847. <https://doi.org/10.1016/j.celrep.2014.03.051> PMID: 24767997
33. Hautakangas MR, Hinttala R, Rantala H, Nieminen P, Uusimaa J, Hassinen IE. Evaluating clinical mitochondrial respiratory chain enzymes from biopsy specimens presenting skewed probability distribution of activity data. *Mitochondrion*. Elsevier B.V. and Mitochondria Research Society; 2016; 29: 53–58. <https://doi.org/10.1016/j.mito.2016.05.004> PMID: 27223842
34. Kramer PA, Chacko BK, George DJ, Zhi D, Wei C-C, Dell'Italia LJ, et al. Decreased Bioenergetic Health Index in monocytes isolated from the pericardial fluid and blood of post-operative cardiac surgery patients. *Biosci Rep*. 2015; 35: e00237–e00237. <https://doi.org/10.1042/BSR20150161> PMID: 26181371
35. Hildick-Smith GJ, Cooney JD, Garone C, Kremer LS, Haack TB, Thon JN, et al. Macrocytic anemia and mitochondriopathy resulting from a defect in sideroflexin 4. *Am J Hum Genet*. The American Society of Human Genetics; 2013; 93: 906–914. <https://doi.org/10.1016/j.ajhg.2013.09.011> PMID: 24119684
36. Pronicka E, Piekutowska-Abramczuk D, Ciara E, Trubicka J, Rokicki D, Karkucińska-Więckowska A, et al. New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. *J Transl Med*. BioMed Central; 2016; 14: 174. <https://doi.org/10.1186/s12967-016-0930-9> PMID: 27290639

37. Haack TB, Rolinski B, Haberberger B, Zimmermann F, Schum J, Strecker V, et al. Homozygous missense mutation in BOLA3 causes multiple mitochondrial dysfunctions syndrome in two siblings. *J Inher Metab Dis*. 2013; 36: 55–62. <https://doi.org/10.1007/s10545-012-9489-7> PMID: 22562699
38. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun*. 2017; 8: 15824. <https://doi.org/10.1038/ncomms15824> PMID: 28604674
39. Van Haute L, Dietmann S, Kremer L, Hussain S, Pearce SF, Powell CA, et al. Deficient methylation and formylation of mt-tRNAMet wobble cytosine in a patient carrying mutations in NSUN3. *Nat Commun*. Nature Publishing Group; 2016; 7: 12039. <https://doi.org/10.1038/ncomms12039> PMID: 27356879
40. Powell CA, Kopajtich R, D'Souza AR, Rorbach J, Kremer LS, Husain RA, et al. TRMT5 Mutations Cause a Defect in Post-transcriptional Modification of Mitochondrial tRNA Associated with Multiple Respiratory-Chain Deficiencies. *Am J Hum Genet*. 2015; 97: 319–328. <https://doi.org/10.1016/j.ajhg.2015.06.011> PMID: 26189817
41. Kremer LS, Distelmaier F, Alhaddad B, Hempel M, Iuso A, Küpper C, et al. Bi-allelic Truncating Mutations in TANGO2 Cause Infancy-Onset Recurrent Metabolic Crises with Encephalocardiomyopathy. *Am J Hum Genet*. 2016; 98: 358–362. <https://doi.org/10.1016/j.ajhg.2015.12.009> PMID: 26805782
42. Seahorse Bioscience. Normalizing XF metabolic data to cellular or mitochondrial parameters, User Guide. 2014; Available: <http://hpst.cz/sites/default/files/attachments/appnote-normalizing-metabolic-data.pdf>
43. Oláhová M, Hardy SA, Hall J, Yarham JW, Haack TB, Wilson WC, et al. LRPPRC mutations cause early-onset multisystem mitochondrial disease outside of the French-Canadian population. *Brain*. 2015; 138: 3503–3519. <https://doi.org/10.1093/brain/awv291> PMID: 26510951
44. Agilent Technologies. Multi-File XF Report Generator, User Guide. 2016; Available: [http://www.agilent.com/cs/library/usermanuals/public/ReportGeneratorUserGuide\\_SeahorseXFCellMitoStressTest\\_MultiFile\\_RevA.pdf](http://www.agilent.com/cs/library/usermanuals/public/ReportGeneratorUserGuide_SeahorseXFCellMitoStressTest_MultiFile_RevA.pdf)