Supplementary Material

**Single cross-validation**

As a general rule, each model was internally validated by single cross-validation as described in Westerhuis et al. (2008). In summary, 4-5 observations were kept out of model development, their tissue TOL concentration was then predicted by the model and subsequently compared with their measured tissue TOL concentration. In this test, the predicted response values ($Q^2$, fraction of the total variation of the response Y that can be predicted) should be close to the measured values ($R^2$). This validation was repeated until all observations were left out once and only once. $P$-values after cross-validation ANOVA were calculated as described in Borel et al. (2014).

$R^2$ and $Q^2$ of the selected model after 100 permutations

This procedure 1) assesses the risk that the selected partial least-square regression model is spurious, i.e. the model fits the current data set well but does not predict Y well for new observations, and 2) tests for overfitting. For this, the accuracy of fit ($R^2$ and $Q^2$) of the original model was compared with the accuracy of fit of 100 models based on data where the order of the Y matrix for the observations (tissue TOL concentration) was randomly permuted, while the X matrix was kept intact. Thus, a robust model (where the fit between X and Y is high) should be unable to predict the permuted Y variables with the intact X variables. Supplementary Figure S1 (below) shows the results of these permutations.
**Supplementary Figure S1** The horizontal axis represents the correlation between the permuted Y’s and the original Y’s. The vertical axis represents the $R^2$ (solid line and triangles) and $Q^2$ (dashed line and squares) values obtained in the permuted models. Values of the original model are on the far right (at correlation = 1), values of the 100 Y-permuted models are further to the left. Note that all the $Q^2$ values (prediction of Y) of the permuted models are lower than the $Q^2$ of the original non permuted model. This strongly supports the conclusion that the ability of the original, non-permuted model to predict the phenotype is not due to chance.
Supplementary Figure S2. Effect of the level of the studied explicative variables on blood TOL concentration using data partitioning

- **n = 32**
  - **mean = 0.72**
  - **SD = 1.27**

- **added VE < 60 mg/kg**
  - **n = 19**
  - **mean = 0.55**
  - **SD = 0.79**

  - **added VE < 20 mg/kg**
    - **n = 4**
    - **mean = 0.29**
    - **SD = 0.36**

  - **added VE ≥ 20 mg/kg**
    - **n = 15**
    - **mean = 0.58**
    - **SD = 0.62**

- **added VE ≥ 60 mg/kg**
  - **n = 13**
  - **mean = 0.88**
  - **SD = 0.75**

  - **MUFA < 35.4%**
    - **n = 9**
    - **mean = 0.83**
    - **SD = 0.61**

  - **MUFA ≥ 35.4%**
    - **n = 4**
    - **mean = 0.99**
    - **SD = 0.12**
The partition algorithm recursively partitions data according to a relationship between the X (explicative variables) and the Y values (TOL concentrations) creating a tree of partitions. It finds a set of cuts or groupings of X values that best predict a Y value. A group of X values is characterized by its mean value (standardized α-tocopherol concentration), its standard deviation (SD) and the number of replication (n).

The 2 explicative variables that significantly explain the variance: added VE: dietary VE dose in diet (in mg/kg); n-6: omega-6 polyunsaturated fatty acids.
Supplementary Figure S3. Effect of the level of the studied explicative variables on liver TOL concentration using data partitioning

- **n = 22**
  - **mean = 0.49**
  - **SD = 1.17**

  - **added VE < 20 mg/kg**
    - **n = 4**
    - **mean = 0.08**
    - **SD = 0.11**
    - **MUFA < 48.6%**
      - **n = 3**
      - **mean = 0.06**
      - **SD = 0.07**
    - **MUFA ≥ 48.6%**
      - **n = 1**
      - **mean = 0.10**
      - **SD = 0.10**

  - **added VE ≥ 20 mg/kg**
    - **n = 18**
    - **mean = 0.61**
    - **SD = 0.88**

  - **added VE < 50 mg/kg**
    - **n = 4**
    - **mean = 0.47**
    - **SD = 0.39**
  - **added VE ≥ 50 mg/kg**
    - **n = 14**
    - **mean = 0.73**
    - **SD = 0.79**

  - **MUFA < 40.2%**
    - **n = 10**
    - **mean = 0.63**
    - **SD = 0.72**
  - **MUFA ≥ 40.2%**
    - **n = 4**
    - **mean = 0.96**
    - **SD = 0.28**
The partition algorithm recursively partitions data according to a relationship between the X (explicative variables) and the Y values (TOL concentrations) creating a tree of partitions. It finds a set of cuts or groupings of X values that best predict a Y value. A group of X values is characterized by its mean value (standardized α-tocopherol concentration), its standard deviation (SD) and the number of replication (n).

The 2 explicative variables that significantly explain the variance: added dietary VE: VE dose in diet (in mg/kg) and MUFA: monounsaturated fatty acids.
**Supplementary Figure S4.** Effect of the level of the studied explicative variables on muscle TOL concentration using data partitioning

- **n = 31**
  - mean = 0.76
  - SD = 1.41

- **added VE < 50 mg/kg**
  - n = 14
  - mean = 0.52
  - SD = 1.11

- **added VE ≥ 50 mg/kg**
  - n = 17
  - mean = 0.94
  - SD = 0.42

  - **added VE < 40 mg/kg**
    - n = 11
    - mean = 0.29
    - SD = 0.43

  - **added VE ≥ 40 mg/kg**
    - n = 3
    - mean = 0.61
    - SD = 1.75

  - **added VE < 193.9 mg/kg**
    - n = 8
    - mean = 0.86
    - SD = 0.36

  - **added VE ≥ 193.9 mg/kg**
    - n = 9
    - mean = 0.97
    - SD = 0.30

  - **n-6 < 51%**
    - n = 3
    - mean = 0.81
    - SD = 0.27

  - **n-6 ≥ 51%**
    - n = 5
    - mean = 0.95
    - SD = 0.34

  - **n-6 < 32%**
    - n = 8
    - mean = 0.34
    - SD = 0.38

  - **n-6 ≥ 32%**
    - n = 5
    - mean = 0.51
    - SD = 0.32

  - **n-6 < 25 mg/kg**
    - n = 3
    - mean = 0.15
    - SD = 0.17

  - **n-6 ≥ 25 mg/kg**
    - n = 8
    - mean = 0.34
    - SD = 0.38
The partition algorithm recursively partitions data according to a relationship between the X (explicative variables) and the Y values (TOL concentrations) creating a tree of partitions. It finds a set of cuts or groupings of X values that best predict a Y value. A group of X values is characterized by its mean value (standardized α-tocopherol concentration), its standard deviation (SD) and the number of replication (n).

The 2 explicative variables that significantly explain the variance: added VE: dietary VE dose in diet (in mg/kg) and n-6: omega-6 polyunsaturated fatty acids.
**Supplementary Figure S5.** Effect of the level of the studied explicative variables on adipose tissue TOL concentration using data partitioning

- **$n = 35$**  
  - **mean = 0.83**  
  - **SD = 1.07**

- **added VE < 40 mg/kg**  
  - $n = 8$  
  - **mean = 0.45**  
  - **SD = 0.67**

- **added VE ≥ 40 mg/kg**  
  - $n = 27$  
  - **mean = 0.89**  
  - **SD = 0.56**

  - **n-6 ≥ 51.5%**  
    - $n = 9$  
    - **mean = 0.85**  
    - **SD = 0.60**

  - **n-6 < 51.5%**  
    - $n = 18$  
    - **mean = 0.93**  
    - **SD = 0.48**

  - **added VE ≥ 112 mg/kg**  
    - $n = 5$  
    - **mean = 0.81**  
    - **SD = 0.40**

  - **added VE < 112 mg/kg**  
    - $n = 4$  
    - **mean = 0.88**  
    - **SD = 0.73**

  - **added VE < 200 mg/kg**  
    - $n = 10$  
    - **mean = 0.89**  
    - **SD = 0.52**

  - **added VE ≥ 200 mg/kg**  
    - $n = 8$  
    - **mean = 0.98**  
    - **SD = 0.26**
The partition algorithm recursively partitions data according to a relationship between the X (explicative variables) and the Y values (TOL concentrations) creating a tree of partitions. It finds a set of cuts or groupings of X values that best predict a Y value. A group of X values is characterized by its mean value (standardized α-tocopherol concentration), its standard deviation (SD) and the number of replication (n).

The 2 explicative variables that significantly explain the variance: added VE: VE dose in diet (in mg/kg) and MUFA: monounsaturated fatty acids.