

## **GENOME-WIDE ASSOCIATION STUDY OF A SEMICONTINUOUS TRAIT: ILLUSTRATION OF THE IMPACT OF THE MODELING STRATEGY THROUGH THE STUDY OF NEUTROPHIL EXTRACELLULAR TRAPS LEVELS**

Gaëlle Munsch<sup>1+</sup>, Carole Proust<sup>1</sup>, Sylvie Labrousche-Colomer<sup>2,3</sup>, Dylan Aïssi<sup>1</sup>, Anne Boland<sup>4</sup>, Pierre-Emmanuel Morange<sup>5,6</sup>, Anne Roche<sup>7</sup>, Luc de Chaisemartin<sup>8,9</sup>, Annie Harroche<sup>10</sup>, Robert Olasso<sup>4,11</sup>, Jean-François Deleuze<sup>4,11</sup>, Chloé James<sup>2,3</sup>, Joseph Emmerich<sup>12,13</sup>, David M Smadja<sup>14,15</sup>, Hélène Jacqmin-Gadda<sup>1\*</sup>, David-Alexandre Tréguët<sup>1\*</sup>

<sup>1</sup> Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, F-33000 Bordeaux, France

<sup>2</sup> UMR1034, Inserm, Biology of Cardiovascular Diseases, University of Bordeaux, Pessac, France

<sup>3</sup> Laboratoire d'Hématologie, CHU de Bordeaux, Pessac, France

<sup>4</sup> Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057, Evry, France

<sup>5</sup> INSERM UMR\_S 1263, Nutrition Obesity and Risk of Thrombosis, Center for CardioVascular and Nutrition research (C2VN), Aix-Marseille University, Marseille 13385, France

<sup>6</sup> Laboratory of Haematology, La Timone Hospital, Marseille 13385, France

<sup>7</sup> Service pneumologie hôpital Bicêtre

<sup>8</sup> Service Auto-immunité, Hypersensibilité et Biothérapies, Hôpital Bichat, Assistance Publique-Hôpitaux de Paris, Paris, France

<sup>9</sup> Université Paris-Saclay, INSERM, Inflammation, Microbiome, Immunosurveillance, Orsay, France

<sup>10</sup> Service d'Hématologie Clinique Centre de Traitement de l'Hémophilie Hôpital Necker Enfants Malades

<sup>11</sup> Centre d'Etude du Polymorphisme Humain, Fondation Jean Dausset, Paris, France

<sup>12</sup> Department of vascular medicine, Paris Saint-Joseph Hospital Group, University of Paris, 185145 rue Raymond Losserand, Paris, 75674, France

<sup>13</sup> UMR1153, INSERM CRESS, 185 rue Raymond Losserand, Paris, 75674, France

<sup>14</sup> Innovative Therapies in Hemostasis, Université de Paris, INSERM, F-75006, Paris, France

<sup>15</sup> Hematology Department and Biosurgical Research Lab (Carpentier Foundation), Assistance Publique Hôpitaux de Paris, Centre-Université de Paris (APHP-CUP), F-75015, Paris, France

\* These authors equally contributed to the work.

+ Corresponding author

## **ABSTRACT**

Semicontinuous data, characterized by an excess of zeros followed by a non-negative and right-skewed distribution, are frequently observed in biomedical research. Different statistical models have been proposed to investigate the association of covariates with such outcome. Motivated by the search of genetic factors associated with Neutrophil Extracellular Traps (NETs), a semicontinuous biomarker involved in thrombosis, we here investigated the impact of the selected model for semicontinuous traits in the context of a Genome Wide Association Study (GWAS). We compared three models that jointly model zero and positive values while allowing the estimation of a single association parameter of covariates with the global mean: Tobit, Negative Binomial and Compound Poisson-Gamma. We assessed the fit of these models to a sample of 657 participants of the FARIVE study measured for NETs plasma levels. For each of these three models, we performed a GWAS on NETs in FARIVE participants and results were compared. A simulation study was also conducted to evaluate the control of the type I error. Compound Poisson-Gamma and Negative Binomial models fitted NETs data observed in FARIVE better than the Tobit model. However, the Negative Binomial model suffered from an inflation of its type I error, attributable to extreme positive values of the NETs and low frequency variants. Conversely, the Compound Poisson-Gamma model was robust to both phenomena. Using the latter model, a GWAS identified a genome wide significant locus on chr21q21.3. The lead variant was rs57502213, a deletion of two nucleotides located ~40kb upstream the non-coding RNA (miR155HG) hosting the miR-155 that was recently highlighted to have a role in NETs formation. This work indicates that the modeling strategy for a semicontinuous outcome in the framework of GWAS studies is crucial. The choice of the model should take into account the nature of the process generating zero values and the presence of extreme values. Our work also suggests that the Compound Poisson-Gamma model, while still marginally employed, can be a robust modeling strategy for GWAS analysis on a semicontinuous trait.

**Key-words:** Semicontinuous outcome / Compound Poisson-Gamma / Negative Binomial / Genome Wide Association Study / Neutrophil Extracellular Traps

## INTRODUCTION

Semicontinuous data, characterized by an excess of zeros followed by a non-negative and right-skewed distribution, are frequently observed in biomedical research<sup>1</sup>. When the study aims at identifying determinants of such a semicontinuous biomarker, it must be handled as the outcome variable and due to the inflation of zeros, classical models such as linear regression cannot be applied without violating the Gaussian assumptions, even with a logarithmic or rank-based inverse-normal transformation. For instance when the interest specifically lies in the identification of molecular determinants associated with a disease semicontinuous biomarker, as it is encountered in the omics era in order to identify/characterize new biological pathways, inform about drug discovery and help in individual risk prediction<sup>2</sup>, the problem of how to model its distribution arises.

Over the past decades several statistical models have been developed to model semicontinuous data by taking into account the mass of zeros. Among the most commonly used models are the Tobit and the two-part models<sup>3,4</sup>.

The two-part model and its extensions<sup>5,6</sup> rely on the use of a logistic regression model to predict the probabilities of occurrence of zero values and of a linear regression model for the analysis of the strictly continuous outcome. The main assumption of this model is that the values of the outcome are derived from two different generating processes. This model has been used in various applications including the modeling of tumor size in cancer, food intake, microbiome abundance or individual costs of chronic kidney disease<sup>7-11</sup>. The two-part model, does not make possible the estimation of a single parameter that represents the association of an explanatory variable on the outcome. In contrast to the two-part model, Tobit models consider a single distribution of the outcome. In the case of zero-inflated data, the Tobit model assumes that the semicontinuous variable is a truncated observation of a Gaussian variable. This modeling is mainly used to account for floor or ceiling effect of the outcome variable that could be due to technical measurement limits<sup>12-15</sup>.

Another possibility is to consider the outcome variable as quantitative discrete, which can be done in some cases by changing the unit of measurement through the use of a multiplicative factor, without losing precision. In this case, models for count data such as the Poisson model or the Negative Binomial model in presence of overdispersion can be used. These models are relevant as long as the proportion of zeros is not too high<sup>16,17</sup>. As the Tobit model, these models allow for a simple interpretation of the results since only one coefficient is estimated per explanatory variable. Extensions of these models have been developed to account for the zero mass (also known as ZIP for Zero-Inflated Poisson and ZINB for Zero-Inflated Negative Binomial) but they make the assumption that the distribution of the outcome is composed of two generating processes, like the two-part models.

New models based on so-called Tweedie distributions<sup>18-20</sup> have recently emerged for the analysis of semicontinuous data but their use remains marginal<sup>21</sup>. The Compound Poisson-Gamma model belongs to this Tweedie family. It assumes that the semicontinuous outcome is defined as a Poisson sum of

gamma random variables. Semicontinuous data are then modeled through the use of a single distribution.

The choice between these different models is not obvious as each semicontinuous outcome has its own properties. As there is no established decision tool, the model to be applied should be chosen according to the distribution of the outcome and the clinical context<sup>22</sup>.

In this work, we show the impact of the adopted model on the results of a Genome Wide Association Study (GWAS) that aimed at identifying genetic factors associated with plasma levels of Neutrophil Extracellular Traps (NETs), a semicontinuous biomarker involved in thrombosis. We highlight the differences between the models with respect to the flexibility of the underlying assumptions, the robustness to outliers and low allele frequency that can help to select the most appropriate model for future studies.

NETs are one of the emerging biomarkers with a key role in thrombosis<sup>23</sup>. In the event of a vascular breach, neutrophils and platelets are the first cells to be recruited and activated<sup>24</sup>. When neutrophils are activated by platelets, they have pro-inflammatory properties that can enhance tissue damage and induce thrombus formation in particular when they evolve towards a certain form of cell death leading to the release of their decondensed chromatin as a network of fibres also called NETs. NETs are composed of DNA fibres comprised of antimicrobial proteins and histones which promote coagulation, platelets activation and thus thrombus formation<sup>25,26</sup>. NETs are involved in many other biological mechanisms such as immune response to viruses, diabetes, cystic fibrosis, cancer tumor growth, progression and metastasis<sup>27-31</sup>.

NETs plasma levels were here measured in 657 participants of the « FActeurs de RIisque et de récidives de la maladie thromboembolique Veineuse » (FARIVE) study<sup>32</sup>. Genome wide genotype data were also available for these participants and then used to conduct a GWAS on NETs levels. We illustrate how the results of this GWAS are impacted by the statistical approach adopted to model NETs plasma levels.

## **METHODS**

### *The FARIVE study*

The FARIVE study is a multicenter case-control study conducted between 2003 and 2007. The sample includes 607 patients with a documented episode of deep vein thrombosis and/or pulmonary embolism and 607 healthy individuals. A detailed description of the study can be found elsewhere<sup>32</sup>. Briefly, patients were not eligible if they were younger than 18 years, had previous venous thrombosis (VT) event, active cancer or recent history of malignancy (within 5 years). Controls were recruited over the same period and matched to cases according to age and sex. They did not have any history of venous and arterial thrombotic disease as well as cancer, liver or kidney failure.

### 1. NETs measurements

Neutrophil Extracellular Traps (NETs) were quantified by measuring myeloperoxidase (MPO)-DNA complexes using an in-house capture ELISA already described<sup>33</sup> in a subsample of 410 VT patients (7 months after their inclusion in the study once the anticoagulant treatment has stopped) and 327 controls (at their time of inclusion in the study). Briefly, microtiter plates were coated with anti-human MPO antibody. After blocking, serum samples were added together with a peroxidase-labeled anti-DNA antibody. After incubation, the peroxidase substrate was added and absorbance measured at 405 nm in a spectrophotometer.

### 2. Genotyping and Imputation

FARIVE participants were genotyped using the Illumina Infinium Global Screening Array v3.0 (GSAv3.0) microarray at the Centre National de Recherche en Génomique Humaine (CNRGH). Individuals with at least one of the following criteria were excluded: discordant sex information, relatedness individuals identified by pairwise clustering of identity by state distances (IBS), genotyping call rate lower than 99%, heterozygosity rate higher/lower than the average rate  $\pm 3$  standard deviation or of non-European ancestry. After applying these criteria, the final sample was composed of 1,077 individuals. Among the 730,059 variants genotyped, 145,238 variants without a valid annotation were excluded as well as 656 variants deviating from Hardy-Weinberg equilibrium in controls at  $P < 10^{-6}$ , 47,286 variants with a Minor Allele Count (MAC) lower than 20 and 1,774 variants with a call rate lower than 95%. This quality controls procedure was conducted using Plink v1.9<sup>34</sup> and the R software v3.6.2. Finally, there were 535,105 markers left for imputation which was then performed with Minimac4 using the 1000 Genomes phase 3 version 5 reference panel<sup>35</sup>.

### 3. Genome Wide Association Study of NETs plasma levels

The present study relies on a subsample of 657 individuals (372 VT cases and 285 controls) with both NETs measurements and imputed genetic data. All single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) greater than 0.01 and imputation quality score greater than 0.3 were tested for association with NETs plasma levels. As shown in the next section, 3 different statistical models were deployed. In all, associations were tested on imputed allelic doses and adjusted for potential confounders that is age, sex, smoking, case-control status and the four first principal components derived from genome wide genotype data<sup>36-38</sup>. The standard genome-wide statistical threshold of  $5 \times 10^{-8}$  was used to consider SNPs as significantly associated with NETs plasma levels.

#### *Statistical modeling for GWAS analysis of NETs plasma levels*

Since we were interested in identifying genetic factors that influence mean NETs plasma levels, any statistical approach that treats independently the zeros mass and the distribution of positive values, such as the two-part, ZIP and ZINB models, was deemed not adapted to our application. As a

consequence, only three models were compared in this study: the Tobit, Compound Poisson-Gamma, and Negative Binomial models. The Poisson model was not investigated because NETs plasma data presented a large overdispersion (see Results section), a situation where Negative Binomial model is preferable.

In this study, we aimed to identify the most suitable model for semicontinuous data in order to conduct a GWAS on NETs and highlight the difference between the three models.

### 1. Tobit model

In the Tobit model, the observed variable  $Y$  is assumed to be a right or left truncated observation of an underlying Gaussian latent variable ( $Y^*$ ). Let  $c$  the constant threshold for truncation which needs to be known and is equal to zero in the context of zero-inflated data. Therefore, the Tobit model assumes that zero values are due to censoring or measurement limits and so they do not represent the true absence of the variable. In the case of a left truncation at 0 the values of the observed variable are:

$$Y = \begin{cases} Y^* & \text{if } Y > 0 \\ 0 & \text{otherwise} \end{cases}$$

The subsequent regression model is:

$$\mathbb{E}(Y^* | X) = \beta X$$

where  $\mathbb{E}(Y^* | X)$  is the expected value of the underlying Gaussian variable  $Y^*$  conditioned on the explanatory variables  $X$ , and where  $\beta$  represents the regression parameters associated to  $X$ . The Tobit model is available in the *VGAM* R package<sup>39</sup>.

### 2. Compound Poisson-Gamma model

An Exponential Dispersion Model (EDM) is a two-parameter family of distributions composed of a linear exponential family with an additional dispersion parameter<sup>40</sup>. EDMs are characterized by their variance function  $\mathbb{V}(\cdot)$  that is an exponential function used to describe the relationship between the mean and the variance. If  $Y$  follows an EDM, then  $\mathbb{E}(Y) = \mu$  and  $Var(Y) = \Phi\mathbb{V}(\mu)$  with  $\Phi$  a dispersion parameter. Tweedie models are a class of EDMs characterized by a power variance function:  $\mathbb{V}(\mu) = \mu^p$  with  $p$  the index parameter<sup>41,42</sup>. Most of the usual distributions are included in the class of Tweedie models such as the normal ( $p = 0$ ), Poisson ( $p = 1$ ), gamma ( $p = 2$ ) and the inverse Gaussian ( $p = 3$ )<sup>43</sup>.

The probability density function of a Tweedie model is defined as<sup>40</sup>:

$$f(y|\mu, \Phi, p) = a(y, \Phi, p) \exp\left(\frac{1}{\Phi} \left( y \frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p} \right)\right)$$

where  $a(\cdot)$  is a given function.

The Compound Poisson-Gamma model belongs to the family of Tweedie models with  $p \in ]1; 2[$ . It simultaneously models the occurrence and the intensity of the semicontinuous outcome<sup>44</sup>. The

distribution of a variable  $Y$  following a Compound Poisson-Gamma model may be defined as a Poisson sum of  $M$  Gamma distributions:

$$Y = \begin{cases} 0, & \text{if } M = 0 \\ K_1 + K_2 + \dots + K_M, & \text{if } M > 0 \end{cases}$$

where  $M \sim Pois(\lambda)$ ,  $K_i \sim Gamma(\alpha, \gamma)$  with  $\alpha$  the shape parameter and  $\gamma$  the scale parameter, and where the values of  $K_i$  are *iid* and independent on  $M$ .

The Compound Poisson-Gamma model is a Tweedie model with the following parametrisation:

$$\mu = \lambda\alpha\gamma; \Phi = \frac{\lambda^{1-p}(\alpha\gamma)^{2-p}}{2-p}; p = \frac{\alpha+2}{\alpha+1} \in ]1; 2[;$$

$$E(Y) = \mu = \lambda\alpha\gamma$$

$$Var(Y) = \Phi\mu^p = \lambda\gamma^2\alpha * (1 + \alpha)$$

Thus, direct modeling of the global expectation  $E(Y)$  is possible using a generalized linear model with a logarithmic link function to insure positivity of the means:

$$\log(E(Y | X)) = \beta X$$

We used the *cplm* R package to implement Compound Poisson-Gamma models<sup>45</sup>.

### 3. Negative Binomial models

We also attempted to use a model for count data by multiplying NETs' values by 1,000 to ensure discreteness without creating new ex-aequos. Let  $Y$  be a random variable following a Poisson distribution which depends on a single parameter  $\lambda > 0$ :

$$E(Y) = Var(Y) = \lambda$$

The Poisson model is adapted to model the expectation of a count variable using a generalized linear model with a logarithmic link function:

$$\log(E(Y | X)) = \beta X$$

The Negative Binomial model is an extension to the Poisson model in the presence of over-dispersion of the outcome:  $Var(Y) > E(Y)$  <sup>46,47</sup>. In that case, the variance of  $Y$  is linked to its expectation through the following relationship:  $Var(Y) = E(Y) + k * E(Y)^2$  where  $k > 0$  is a dispersion parameter. This model is also part of generalized linear models and its link function is the logarithm. Negative Binomial model can also be represented as Poisson distributions with a Gamma distributed means where  $Y \sim Pois(\lambda)$  and  $\lambda \sim Gamma(\alpha, \gamma)$  <sup>48</sup>. However, unlike the Compound Poisson-Gamma presented above, the two variables  $Y$  and  $\lambda$  are not independent from each other.

#### *Models' comparison*

The three aforementioned tested models were applied to NETs data while adjusting for age, sex, smoking and case-control status.

The fit of these models to NETs data were assessed in two ways. First, we computed the Root Mean Square Error (RMSE) of each tested model defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

where  $N$  is the sample size,  $\hat{y}_i$  the prediction of the  $i^{th}$  individual according to its covariates provided by a given modeling approach and  $y_i$  the observed value. Instead of predicting  $\hat{y}_i$  by  $\mathbb{E}(Y | X, \hat{\beta})$  that cannot be equal to zero, we used simulated predictions. For each studied model, a random value was generated for each individual according to its explanatory variables and the estimated model parameters. This process was repeated 1,000 times and the mean of RMSEs over the 1000 replicates was reported. As the Tobit model predicts negative values that are not observed in our semicontinuous outcome, these were imputed at zero to calculate the corresponding RMSE.

Second, we graphically assessed the fit of models predictions using Quantile-Quantile plot (QQplot) of the observations and predictions for each tested model.

### *Simulation Study*

A simulation study was conducted to evaluate the control of the type I error ( $\alpha$ ) of the Negative Binomial and Compound Poisson-Gamma models in the context of genetic association studies as well as their sensibility to outliers. From the observed NETs data distribution, we randomly generated  $S = 1, \dots, 10000$  bootstrapped samples of size  $N = 657$ . For each bootstrapped sample, all individuals were randomly assigned 4 independent genotypes under the assumption of Hardy-Weinberg and corresponding to 4 SNPs with allele frequencies 0.01, 0.05, 0.10 and 0.20. The association of SNPs with the outcome was tested under the assumption of additive allele effects. This procedure was used to simulate semicontinuous data that mimic the NETs data observed in FARIVE and to allow the evaluation of the robustness of the two studied models (Negative Binomial and Compound Poisson-Gamma models) to a deviation from their underlying distribution. To assess the robustness to outliers, each simulated dataset was also analysed after the exclusion of individuals with simulated trait higher than 0.5, a threshold corresponding on average to the exclusion of 3% of individuals with the highest NETs levels.

For each tested model, the number of times a SNP was found statistically significant at  $\alpha=0.05$ ,  $\alpha=0.01$ , and  $\alpha=0.001$  was used to compute its empirical type I error.



## RESULTS

### *Population characteristics*

The main characteristics of the FARIVE participants used in this work are presented in **Table 1**. There is approximately 40% of men, 20% of current smokers and individuals are on average 53 years old. The distribution of NETs plasma levels observed in FARIVE is shown in **Figure 1A**. Approximately 16% of exact zeros were observed with a higher proportion among VT cases compared to controls (20% and 10% respectively). To analyse NETs as count data, observed values were multiplied by 1,000. This induced a large overdispersion (mean=78; variance=26,064) leading to the adoption of a Negative Binomial model for analysing such data.

### *Clinical Variables & Goodness of fit*

**Table 2** reports the association of clinical covariates with NETs plasma levels in each of the three studied model. The Tobit model assumes a linear association of the covariates on the expected mean of the latent variable, i.e. the true value of NETs. For example, each 10-year increase in age is associated with an increase of 0.005 on the expected mean of the latent variable of NETs plasma levels, given the other covariates are held constant. Regarding the two other models, as a logarithmic link function is used, the association of covariates on the expected mean of NETs is multiplicative. As a consequence, for the Compound Poisson-Gamma model, each 10-year increase of age is associated with an expected mean of NETs plasma levels multiplied by 1.05 ( $= e^{0.05}$ ). Similar interpretation holds for the Negative Binomial model that yielded regression parameters very close to those obtained via the Compound Poisson-Gamma models.

RMSEs provided by the three models are shown in **Table 2**. The lowest RMSE was observed for the Compound Poisson-Gamma model while the Negative Binomial model exhibited the highest one. Graphically, the Compound Poisson-Gamma (**Figure 1C**) and the Negative Binomial (**Figure 1D**) showed similar distributions of their predicted values even if the right skewedness was slightly less pronounced for the Compound Poisson-Gamma distribution. These distributions were rather close to that observed for the original NETs data (**Figure 1A**). By contrast, the Tobit distribution (**Figure 1B**) substantially deviated from the original data and looked like a left-truncated Gaussian distribution.

Quantile-Quantile plots of the observed vs predicted values did not visually show obvious deviation from the bisection line, except for high values (above 0.5 in the original NET scale), for the Compound Poisson-Gamma (**Figure 2B**) and the Negative Binomial (**Figure 2C**) models. Conversely, for the Tobit model (**Figure 2A**), the QQplot line deviated from the bisection line from the lower values.

Altogether, these observations suggest that the Compound Poisson-Gamma model seems the most adequate to analyze FARIVE NETs data. Nevertheless, we conducted a GWAS on NETs plasma

levels using each of the three models discussed above in order to get additional elements of comparison between these models.

### *GWAS analysis on NETs plasma levels*

A total of 9,670,724 autosomal SNPs were tested for association with NETs plasma levels using the Tobit, Negative Binomial and Compound Poisson-Gamma models. Quantile-Quantile plots for the observed and expected p-values summarizing the GWAS results for each model are shown in **Figure 3**. While the whole set of association results was compatible with what was expected under the null hypothesis of no genetic association for the Compound Poisson-Gamma model (**Figure 3B**), strong deviations were observed for the Tobit (**Figure 3A**) and Negative Binomial (**Figure 3C**) models. By restricting the GWAS results to SNPs with MAF greater than 5%, inflation was no longer observed for the Tobit model (**Supplementary Figure S1A**, genomic inflation factor  $\lambda=0.96$ ) while the Negative Binomial model remained strongly inflated (**Supplementary Figure S1C**,  $\lambda=1.46$ ).

To further explore the remaining inflation, we re-ran the GWAS under the Compound Poisson-Gamma and Negative Binomial models after excluding 19 FARIVE participants (~3%) with NETs plasma levels higher than 0.5. Inflation in the Negative Binomial model was considerably decreased (**Supplementary Figure S2B**) and completely vanished when we additionally restricted the GWAS analysis on SNP with MAF >5% (**Supplementary Figure S3B**,  $\lambda=1.03$ ). Finally, the simulation studies demonstrated that the test of SNP in the Negative Binomial model exhibited inflated type I error for  $\alpha$  for the three nominal values of alpha considered (**Supplementary Table S1**) when data distribution fit the one here observed for NETs plasma levels in the FARIVE study. Type I error was rather well controlled in absence of extreme values (**Supplementary Table S2**). Of note, these simulations also show that the Compound Poisson-Gamma model generally controls the nominal type I error. All these observations add support for the use of the Compound Poisson-Gamma model for the GWAS analysis of NETs plasma levels.

The corresponding Manhattan plot shown in **Figure 4** revealed one genome-wide significant locus. The lead SNP rs57502213 is a deletion of two nucleotides (TC), mapping to the miR-155 hosting gene (MIR155HG). This variant had a MAF of ~7%, exhibited a good imputation quality ( $R_{sq}=0.92$ ) and its minor allele was associated with a 2.53 fold increase (95% confidence interval [1.85 – 3.47],  $P=1.42 \times 10^{-8}$ ) in NETs plasma levels. The average NETs plasma levels were higher in carriers of the deletion of the TC allele than in non-carriers (0.15 vs 0.07). This pattern of association was consistent in VT cases and in controls (**Table 4**). Full GWAS summary statistics will be available on GWAS catalog<sup>49</sup>.

## **DISCUSSION**

This work was motivated by the search of genetic factors associated with NETs plasma levels exhibiting a semicontinuous distribution. We compared three different modeling strategies, Tobit,

Negative Binomial and Compound Poisson-Gamma models, that handle the excess of zero and the asymmetric distribution while allowing the estimation of a single regression parameter for characterizing the association between an explanatory variable and the global mean of the semicontinuous outcome.

Visual inspection showed that both the Negative Binomial and Compound Poisson-Gamma models fit better the observed NETs distribution than the Tobit model. Indeed, the underlying hypothesis of a left-truncated Gaussian distribution with only two parameters makes the Tobit model less-flexible than Compound Poisson-Gamma and Negative Binomial models. RMSE analysis provided further support for the use of Compound Poisson-Gamma model. Of note, the definition of this model matches quite well the biological mechanisms underlying NETs production as it is intuitively reasonable to assume that the number of dead neutrophils follows a Poisson distribution, and that each of these rejects a certain quantity of NETs that would follow a Gamma distribution.

Our GWAS and simulation studies revealed that the Tobit and Negative Binomial models were prone to strong inflation of p-values. While this inflation could be attributable to SNPs with low allele frequency (MAF<5%) for the Tobit model, this inflation was due to both low allele frequency SNPs and extreme positive values of the outcome for the Negative Binomial model. The Compound Poisson-Gamma model was much more robust to these two phenomena.

The Compound Poisson-Gamma based GWAS identified one significant locus on chr21q21.3 associated with NETs plasma levels. This locus maps to miR155HG and the lead SNP was rs57502213, an intronic deletion in miR155HG. While several recent publications have highlighted the role of the associated miRNA in the NETs formation<sup>50-52</sup>, little information is available in public resources about the possible functional impact of rs57502213. This SNP is in moderate linkage disequilibrium (pairwise  $D' > 0.50$ ) with several other nearby variants located in *MRPL39*, *GABPA* and *APP*, the latter having also been reported to be involved in NETs formation<sup>53</sup>. Note that another GWAS on NETs plasma levels has recently been conducted in the Rotterdam study<sup>54</sup>. Despite the large sample size of this study (~5600 individuals), no significant genome-wide association was detected and the association of our lead SNP did not replicate there ( $P=0.14$ ). However, different kits were used to measure NETs levels in the two studies and recent work has emphasized the need for standardized methods for NETs measurements<sup>55,56</sup> ([https://cdn.ymaws.com/www.isth.org/resource/resmgr/ssc/ssc\\_subcommittee\\_project\\_mar.pdf](https://cdn.ymaws.com/www.isth.org/resource/resmgr/ssc/ssc_subcommittee_project_mar.pdf)). Of note, in the Rotterdam study, NETs were analyzed using a log-transformed model suggesting that no zero values (or little) were observed. This contrasts with FARIVE data and could contribute to the heterogeneity of findings between studies. Nevertheless, because of the increasingly recognized role of the chr21q21.3 locus in NETs biology, further works deserve to be conducted to clarify the genetic signal observed in the present study.

To conclude, our work indicates that the modeling strategy for a semicontinuous outcome is crucial, but not straightforward. The choice of the model should take into account the nature of the (biological) process generating zero values, the distribution of the outcome and, especially, the presence of extreme values. The Tobit model with only two parameters is less flexible than Compound Poisson-Gamma and Negative Binomial models and our work suggests that the Compound Poisson-Gamma model, while still marginally used, is more robust than the Negative Binomial model to outliers and low allele frequency making. This make it well suitable for GWAS analysis on semicontinuous trait.

## **STATEMENTS AND DECLARATIONS**

### *Acknowledgments*

GM benefited from the EUR DPH, a PhD program supported within the framework of the PIA3 (Investment for the future). Project reference 17-EURE-0019.

Statistical analyses benefited from the CBiB computing centre of the University of Bordeaux.

This project was carried out in the framework of the INSERM GOLD Cross-Cutting program (P-EM, D-AT).

### *Fundings*

GM and D-AT are supported by the EPIDEMIOIOM-VT Senior Chair from the University of Bordeaux initiative of excellence IdEX.

The FARIVE study was supported by grants from the Fondation pour la Recherche Médicale, the Program Hospitalier de recherche Clinique (PHRC 20 002; PHRC2009 RENOVATV), the Fondation de France, and the Leducq Foundation. FARIVE genetic data were funded by the GENMED Laboratory of Excellence on Medical Genomics [ANR-10-LABX-0013], a research program managed by the National Research Agency (ANR) as part of the French Investment for the Future.

### *Disclosures*

The authors have no conflict of interest to declare.

### *Ethics approval*

Research have been performed in accordance with the Declaration of Helsinki. The FARIVE study was approved by the “Comité consultatif de protection des personnes dans la recherche biomédicale” (Project n° 2002-034).

### *Consent to participate*

Written informed consent to participate was obtained from all FARIVE participants.

## REFERENCES

1. Min Y, Agresti A. Modeling Nonnegative Data with Clumping at Zero: A Survey. :27.
2. Folkersen L, Gustafsson S, Wang Q, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab.* 2020;2(10):1135-1148. doi:10.1038/s42255-020-00287-2
3. Cragg JG. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica.* 1971;39(5):829-844. doi:10.2307/1909582
4. Tobin J. Estimation of Relationships for Limited Dependent Variables. *Econometrica.* 1958;26(1):24-36. doi:10.2307/1907382
5. Farewell VT, Long DL, Tom BDM, Yiu S, Su L. Two-Part and Related Regression Models for Longitudinal Data. *Annu Rev Stat Its Appl.* 2017;4:283-315. doi:10.1146/annurev-statistics-060116-054131
6. Feng X, Lu B, Song X, Ma S. Financial literacy and household finances: A Bayesian two-part latent variable modeling approach. *J Empir Finance.* 2019;51:119-137. doi:10.1016/j.jempfin.2019.02.002
7. Chen EZ, Li H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics.* 2016;32(17):2611-2617. doi:10.1093/bioinformatics/btw308
8. Garbutt DJ, Stern RD, Dennett MD, Elston J. A comparison of the rainfall climate of eleven places in West Africa using a two-part model for daily rainfall. *Arch Meteorol Geophys Bioclimatol Ser B.* 1981;29(1):137-155. doi:10.1007/BF02278197
9. Hartman B, Larson C, Kunkel C, Wight C, Warr RL. A Two-Part Model of the Individual Costs of Chronic Kidney Disease. :60.
10. Rustand D, Briollais L, Tournigand C, Rondeau V. Two-part joint model for a longitudinal semicontinuous marker and a terminal event with application to metastatic colorectal cancer data. *Biostatistics.* 2022;23(1):50-68. doi:10.1093/biostatistics/kxaa012
11. Tooze JA, Midthune D, Dodd KW, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc.* 2006;106(10):1575-1587. doi:10.1016/j.jada.2006.07.003
12. Amore MD, Murtinu S. Tobit models in strategy research: Critical issues and applications. *Glob Strategy J.* 2021;11(3):331-355. doi:10.1002/gsj.1363
13. Chen T, Ma S, Kobie J, Rosenberg A, Sanz I, Liang H. Identification of significant B cell associations with undetected observations using a Tobit model. *Stat Interface.* 2016;9(1):79-91. doi:10.4310/SII.2016.v9.n1.a8
14. Debnath AK, Blackman R, Haworth N. A Tobit model for analyzing speed limit compliance in work zones. *Saf Sci.* 2014;70:367-377. doi:10.1016/j.ssci.2014.07.012
15. McBee M. Modeling Outcomes With Floor or Ceiling Effects: An Introduction to the Tobit Model. *Gift Child Q.* 2010;54(4):314-320. doi:10.1177/0016986210379095
16. van den Broek J. A Score Test for Zero Inflation in a Poisson Distribution. *Biometrics.* 1995;51(2):738-743. doi:10.2307/2532959

17. Allison PD. *Logistic Regression Using SAS: Theory and Application, Second Edition*. SAS Institute; 2012.
18. Tweedie MC. An index which distinguishes between some important exponential families. In: Vol 579. ; 1984:579-604.
19. Gilchrist R, Drinkwater D. The use of the Tweedie distribution in statistical modelling. In: Bethlehem JG, van der Heijden PGM, eds. *COMPSTAT*. Physica-Verlag HD; 2000:313-318. doi:10.1007/978-3-642-57678-2\_39
20. Jørgensen B, Martínez JR, Vinogradov V. Domains of attraction to Tweedie distributions. *Lith Math J*. 2009;49(4):399-425. doi:10.1007/s10986-009-9062-8
21. Kurz CF. Tweedie distributions for fitting semicontinuous health care utilization cost data. *BMC Med Res Methodol*. 2017;17(1):171. doi:10.1186/s12874-017-0445-y
22. Brown JE, Dunn PK. Comparisons of Tobit, Linear, and Poisson-Gamma Regression Models: An Application of Time Use Data. *Sociol Methods Res*. 2011;40(3):511-535. doi:10.1177/0049124111415370
23. Kimball AS, Obi AT, Diaz JA, Henke PK. The Emerging Role of NETs in Venous Thrombosis and Immunothrombosis. *Front Immunol*. 2016;7:236. doi:10.3389/fimmu.2016.00236
24. Ruhnau J, Schulze J, Dressel A, Vogelgesang A. Thrombosis, Neuroinflammation, and Poststroke Infection: The Multifaceted Role of Neutrophils in Stroke. *J Immunol Res*. 2017;2017:5140679. doi:10.1155/2017/5140679
25. Laridan E, Martinod K, De Meyer SF. Neutrophil Extracellular Traps in Arterial and Venous Thrombosis. *Semin Thromb Hemost*. 2019;45(1):86-93. doi:10.1055/s-0038-1677040
26. Diamond SL. Systems Analysis of Thrombus Formation. *Circ Res*. 2016;118(9):1348-1362. doi:10.1161/CIRCRESAHA.115.306824
27. Ng H, Havervall S, Rosell A, et al. Circulating Markers of Neutrophil Extracellular Traps Are of Prognostic Value in Patients With COVID-19. *Arterioscler Thromb Vasc Biol*. 2021;41(2):988-994. doi:10.1161/ATVBAHA.120.315267
28. Wang L, Zhou X, Yin Y, Mai Y, Wang D, Zhang X. Hyperglycemia Induces Neutrophil Extracellular Traps Formation Through an NADPH Oxidase-Dependent Pathway in Diabetic Retinopathy. *Front Immunol*. 2019;9:3076. doi:10.3389/fimmu.2018.03076
29. Zhu L, Liu L, Zhang Y, et al. High Level of Neutrophil Extracellular Traps Correlates With Poor Prognosis of Severe Influenza A Infection. *J Infect Dis*. 2018;217(3):428-437. doi:10.1093/infdis/jix475
30. Martínez-Alemán SR, Campos-García L, Palma-Nicolas JP, Hernández-Bello R, González GM, Sánchez-González A. Understanding the Entanglement: Neutrophil Extracellular Traps (NETs) in Cystic Fibrosis. *Front Cell Infect Microbiol*. 2017;7:104. doi:10.3389/fcimb.2017.00104
31. Masucci MT, Minopoli M, Del Vecchio S, Carriero MV. The Emerging Role of Neutrophil Extracellular Traps (NETs) in Tumor Progression and Metastasis. *Front Immunol*. 2020;11:1749. doi:10.3389/fimmu.2020.01749
32. Trégouët DA, Heath S, Saut N, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood*. 2009;113(21):5298-5303. doi:10.1182/blood-2008-11-190389

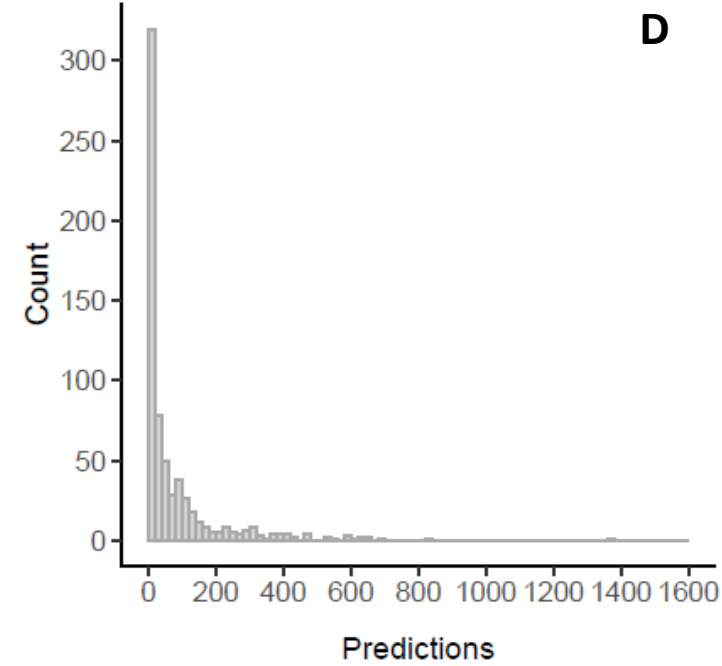
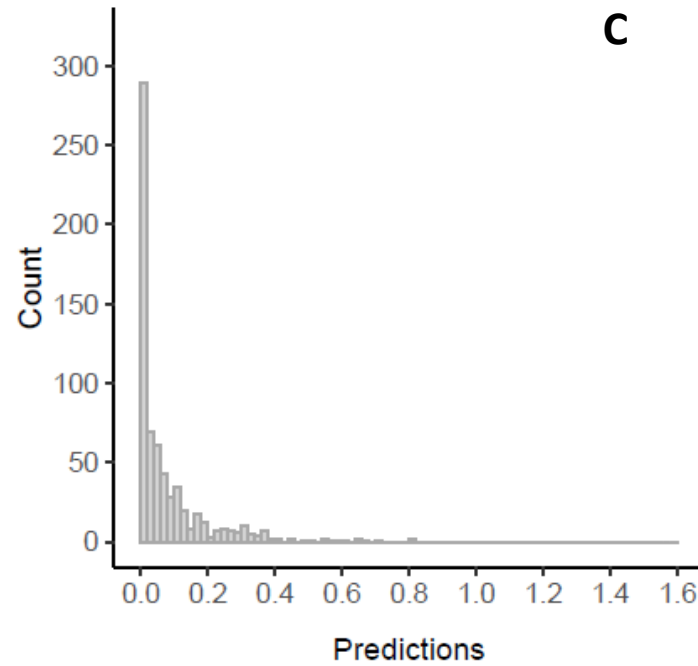
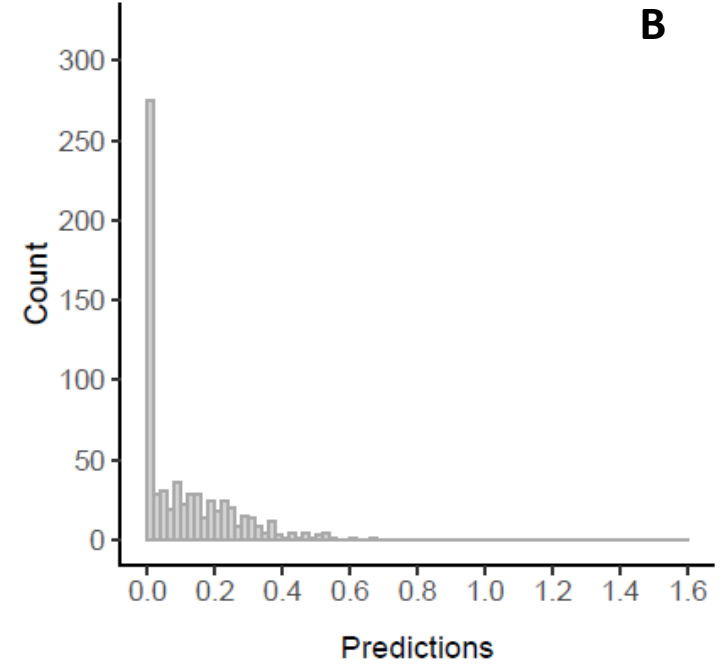
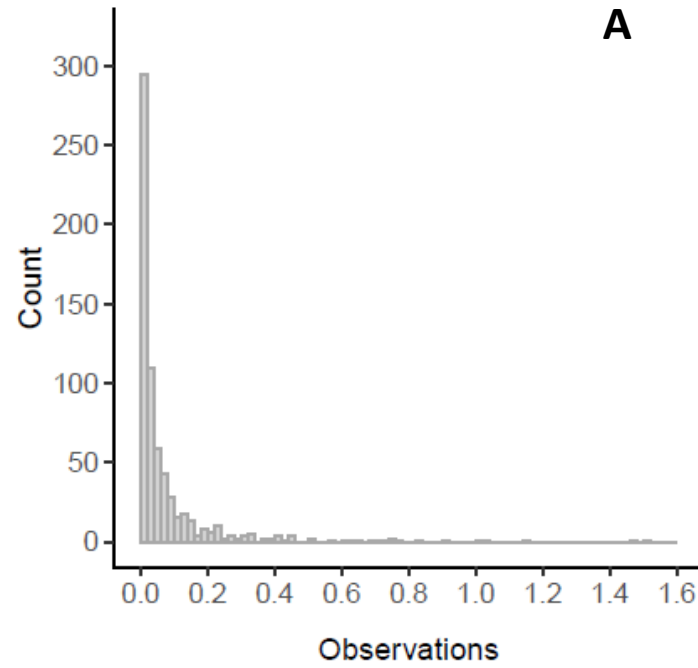
33. Granger V, Peyneau M, Chollet-Martin S, de Chaisemartin L. Neutrophil Extracellular Traps in Autoimmunity and Allergy: Immune Complexes at Work. *Front Immunol.* 2019;10:2824. doi:10.3389/fimmu.2019.02824
34. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4(1):7. doi:10.1186/s13742-015-0047-8
35. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284-1287. doi:10.1038/ng.3656
36. White PC, Hirschfeld J, Milward MR, et al. Cigarette smoke modifies neutrophil chemotaxis, neutrophil extracellular trap formation and inflammatory response-related gene expression. *J Periodontal Res.* 2018;53(4):525-535. doi:10.1111/jre.12542
37. Ortmann W, Kolaczowska E. Age is the work of art? Impact of neutrophil and organism age on neutrophil extracellular trap formation. *Cell Tissue Res.* 2018;371(3):473-488. doi:10.1007/s00441-017-2751-4
38. Yuan XZ. Sex differences in neutrophil extracellular trap formation. :116.
39. Yee TW. *Vector Generalized Linear and Additive Models: With an Implementation in R.* Springer; 2015.
40. Zhou H, Qian W, Yang Y. Tweedie gradient boosting for extremely unbalanced zero-inflated data. *Commun Stat - Simul Comput.* Published online July 11, 2020:1-23. doi:10.1080/03610918.2020.1772302
41. Jørgensen B. Exponential Dispersion Models. *J R Stat Soc Ser B Methodol.* 1987;49(2):127-145. doi:10.1111/j.2517-6161.1987.tb01685.x
42. Fox J. *Applied Regression Analysis and Generalized Linear Models.* Third edition. SAGE; 2016.
43. Dunn PK, Smyth GK. Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Stat Comput.* 2008;18(1):73-86. doi:10.1007/s11222-007-9039-6
44. Dzupire NC, Ngare P, Odongo L. A Poisson-Gamma Model for Zero Inflated Rainfall Data. *J Probab Stat.* 2018;2018:1-12. doi:10.1155/2018/1012647
45. Zhang Y. Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Stat Comput.* 2013;23(6):743-757. doi:10.1007/s11222-012-9343-7
46. Hoef JMV, Boveng PL. Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data? *Ecology.* 2007;88(11):2766-2772.
47. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol Bull.* 1995;118(3):392-404. doi:10.1037/0033-2909.118.3.392
48. Gorshenin AK, Korolev VYu. Scale Mixtures of Frechet Distributions as Asymptotic Approximations of Extreme Precipitation. *J Math Sci.* 2018;234(6):886-903. doi:10.1007/s10958-018-4052-1
49. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120

50. Saha P, Yeoh BS, Xiao X, et al. PAD4-dependent NETs generation are indispensable for intestinal clearance of *Citrobacter rodentium*. *Mucosal Immunol.* 2019;12(3):761-771. doi:10.1038/s41385-019-0139-3
51. Hawez A, Taha D, Algaber A, Madhi R, Rahman M, Thorlacius H. MiR-155 regulates neutrophil extracellular trap formation and lung injury in abdominal sepsis. *J Leukoc Biol.* 2022;111(2):391-400. doi:10.1002/JLB.3A1220-789RR
52. Rohrbach A, Slade D, Thompson P, Mowen K. Activation of PAD4 in NET formation. *Front Immunol.* 2012;3. Accessed August 1, 2022. <https://www.frontiersin.org/articles/10.3389/fimmu.2012.00360>
53. Canobbio I, Visconte C, Momi S, et al. Platelet amyloid precursor protein is a modulator of venous thromboembolism in mice. *Blood.* 2017;130(4):527-536. doi:10.1182/blood-2017-01-764910
54. Donkel SJ, Portilla Fernández E, Ahmad S, et al. Common and Rare Variants Genetic Association Analysis of Circulating Neutrophil Extracellular Traps. *Front Immunol.* 2021;12:615527. doi:10.3389/fimmu.2021.615527
55. Prével R, Dupont A, Labrousche-Colomer S, et al. Plasma Markers of Neutrophil Extracellular Trap Are Linked to Survival but Not to Pulmonary Embolism in COVID-19-Related ARDS Patients. *Front Immunol.* 2022;13:851497. doi:10.3389/fimmu.2022.851497
56. Rada B. Neutrophil Extracellular Traps. *Methods Mol Biol Clifton NJ.* 2019;1982:517-528. doi:10.1007/978-1-4939-9424-3\_31



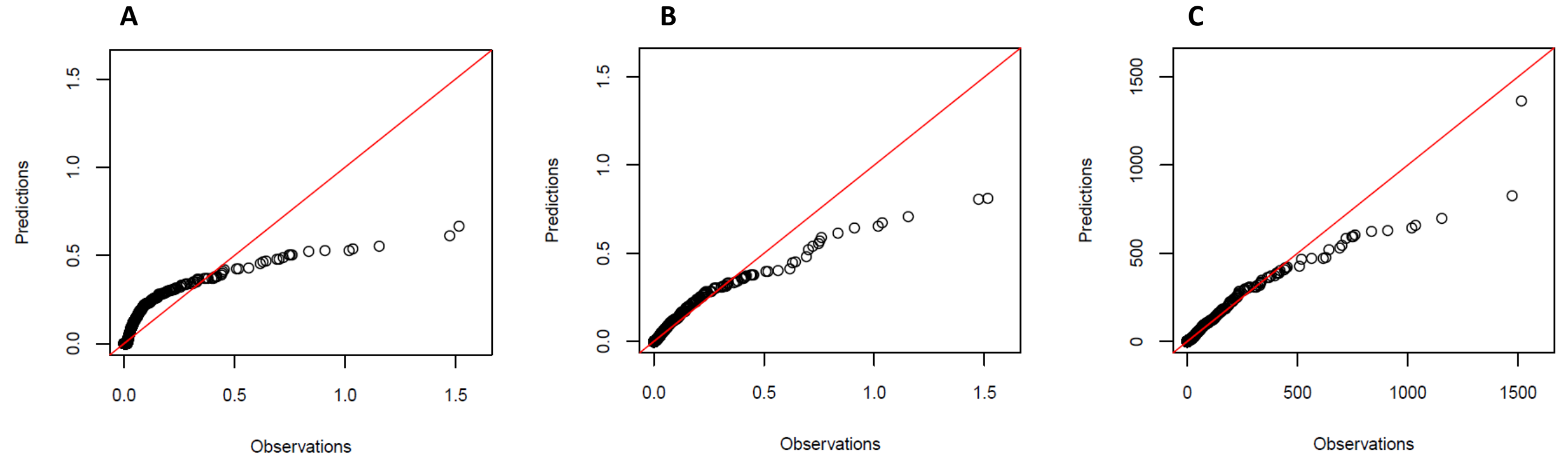
### Figure 1 : Distribution of NETs plasma levels

This figure presents the distribution of the observed NETs plasma levels (A), predictions from the Tobit model (B), the Compound Poisson-Gamma model (C) and the Negative Binomial model (D).



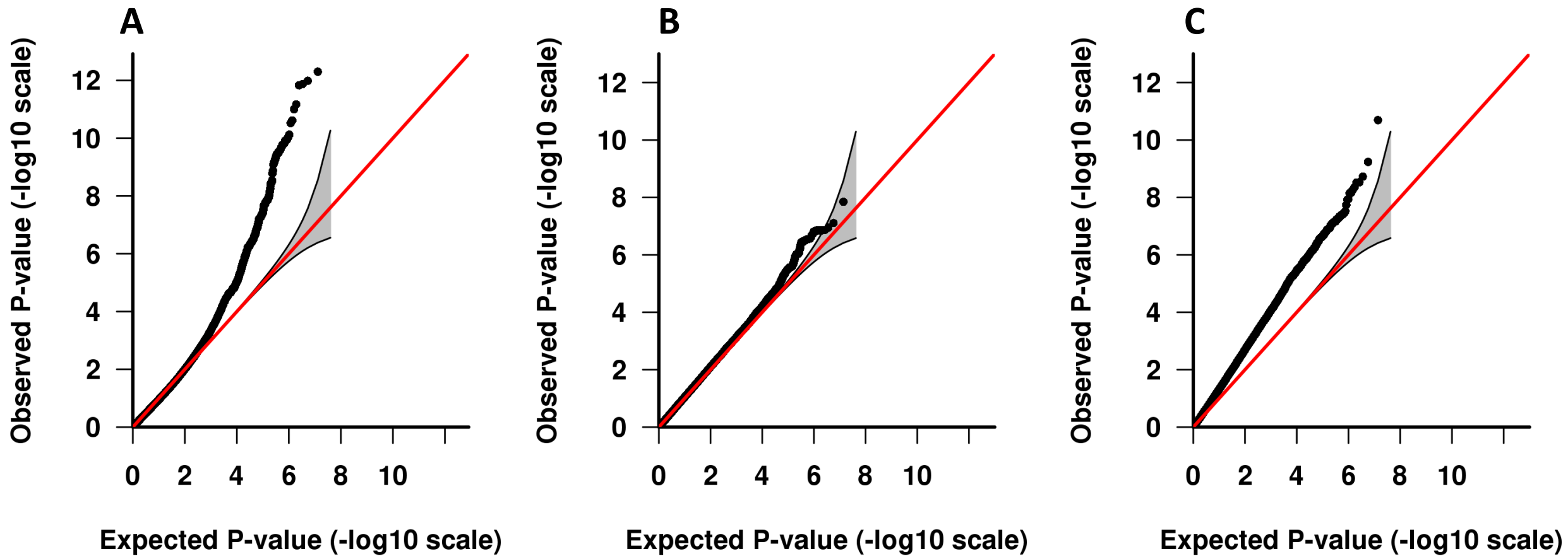
## Figure 2 : QQplots of observations and predictions from the three models

From Tobit (A), Compound Poisson-Gamma (B) and Negative Binomial (C).



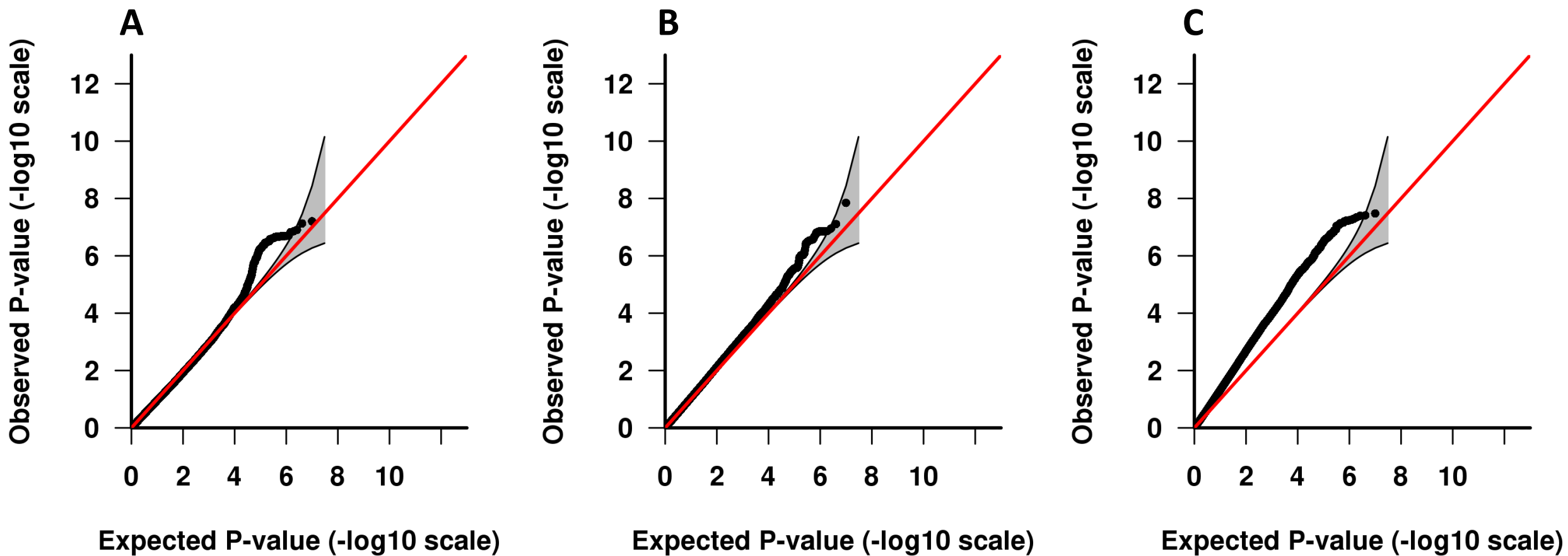
**Figure 3: Quantile-Quantile plots from the GWAS results on NETs plasma levels**

Tobit (A), Compound Poisson-Gamma (B) and Negative Binomial (C).



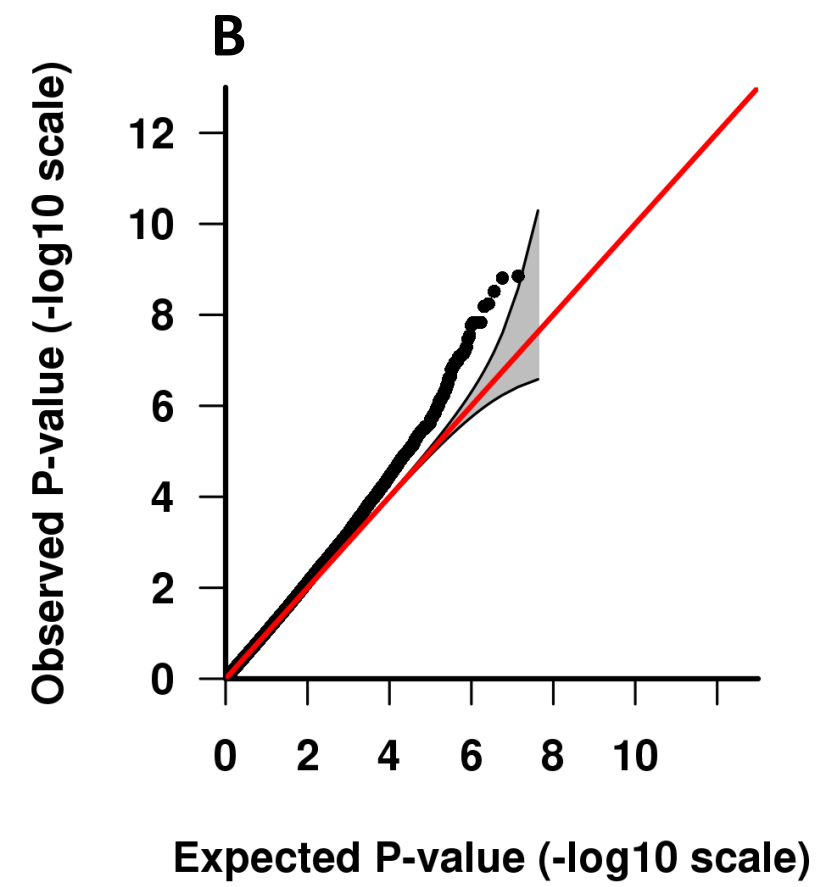
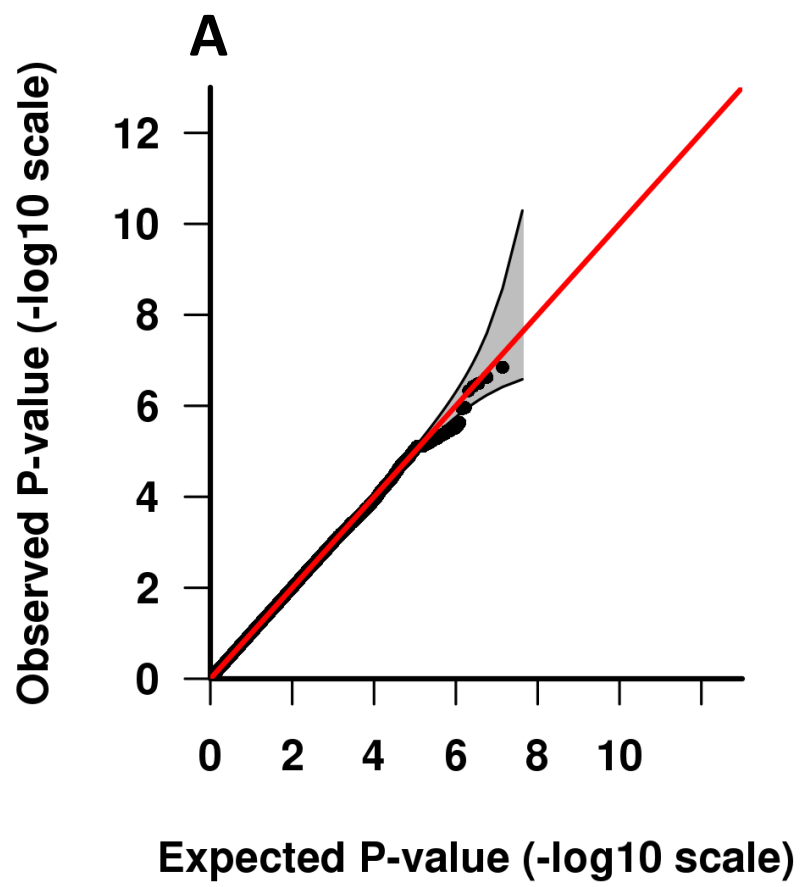
# Supplementary Figure S1: Quantile-Quantile plots from the GWAS results with an allele frequency higher than 5% on NETs plasma levels

Tobit (A), Compound Poisson-Gamma (B) and Negative Binomial (C).



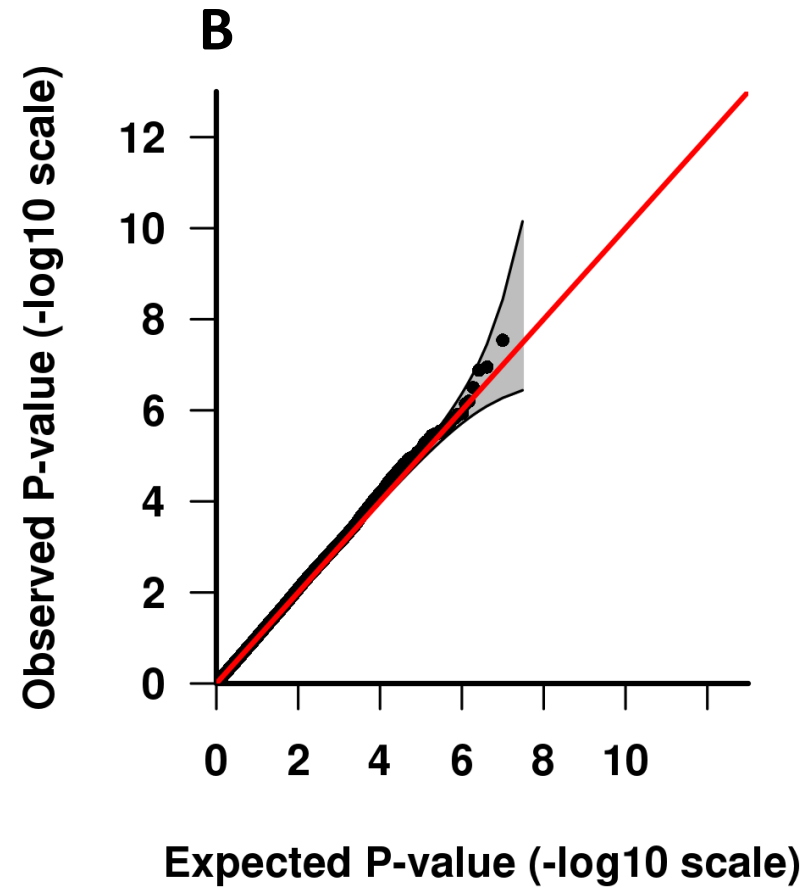
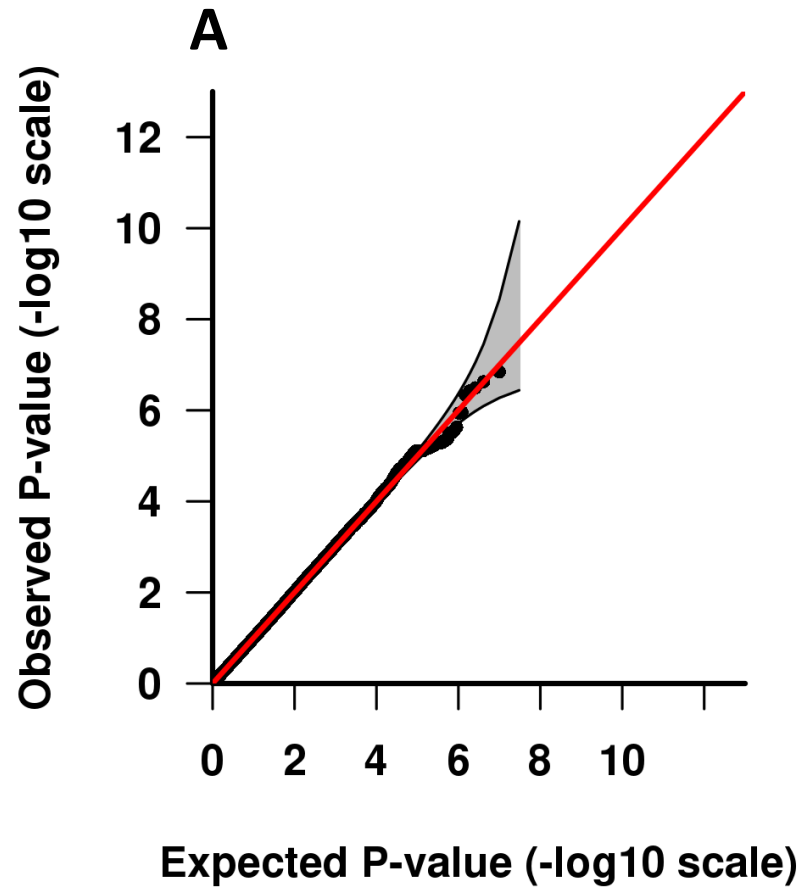
# Supplementary Figure S2 : Quantile-Quantile plots from the GWAS results on NETs plasma levels without outliers

Compound Poisson-Gamma (A) and Negative Binomial (B).

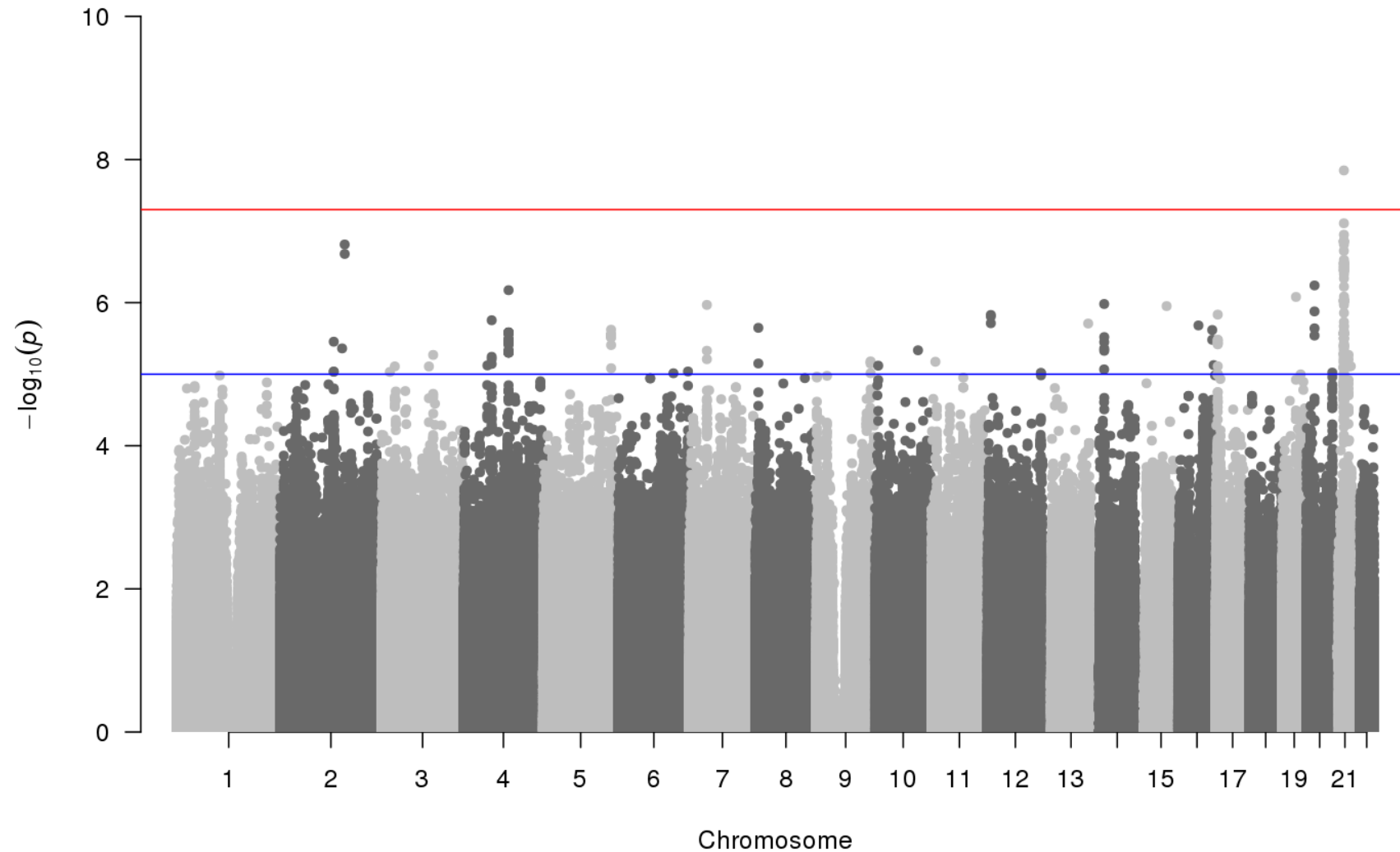


# Supplementary Figure S3: Quantile-Quantile plots from the GWAS results with an allele frequency higher than 5% on NETs plasma levels without outliers

Compound Poisson-Gamma (A) and Negative Binomial (B).



**Figure 4: Manhattan plots from the GWAS results of NETs conducted with Compound Poisson-Gamma model**



The  $-\log_{10}$  of the p-values are presented according to the position of the associated tested SNP across the genome. The genome wide significant threshold ( $5 \times 10^{-8}$ ) is represented with a red line.

<b>Table 1: Main characteristics of the FARIVE study.....</b>	<b>2</b>
<b>Table 2: Comparison of regression parameter estimates according to the three studied modeling .....</b>	<b>3</b>
<b>Table 3 : Association rs57502213 of with NETs plasma levels.....</b>	<b>4</b>
<b>Supplementary Table S1 : Evaluation of the control of the type I error according to the frequency of the SNP tested with Compound Poisson-Gamma and Negative Binomial models....</b>	<b>5</b>
<b>Supplementary Table S2: Evaluation of the control of the type I error when outliers are removed, according to the frequency of the SNP tested with Compound Poisson-Gamma and Negative Binomial models .....</b>	<b>5</b>



**Table 1 Main characteristics of the FARIVE study**

	<b>Total N=657</b>	<b>VT‡ Cases N=372</b>	<b>Controls N=285</b>
	<b>N (%)</b>	<b>N (%)</b>	<b>N (%)</b>
<b>Sex - Men</b>	256 (39.0%)	141 (37.9%)	115 (40.4%)
<b>Age at sampling</b> ( <i>Mean ± SD</i> )	53.0 ± 18.8	53.3 ± 19.3	52.8 ± 18.2
<b>Smoking status</b>			
Current smoker	128 (19.5%)	62 (16.7%)	66 (23.2%)
Former smoker / Never	529 (80.5%)	310 (83.3%)	219 (76.8%)
<b>Neutrophil Extracellular Traps levels</b>			
All values			
Mean ± SD	0.08 ± 0.16	0.05 ± 0.11	0.12 ± 0.20
Median [Q1;Q3]	0.03 [4x10 <sup>-3</sup> ;0.07]	0.03 [2x10 <sup>-3</sup> ;0.05]	0.04 [0.01;0.12]
Exact zero	104 (15.8%)	75 (20.2%)	29 (10.2%)

□ Standard Deviation

‡ Venous Thrombosis

**Table 2: Comparison of regression parameter estimates according to the three models**

	<b>Tobit</b>	<b>Compound Poisson-Gamma</b>	<b>Negative Binomial*</b>
	Beta (SD)	Beta (SD)	Beta (SD)
<i>Covariates</i>			
<b>Age (10y)</b>	0.005 (0.004)	0.05 (0.04)	0.05 (0.04)
<b>Sex (Males)</b>	-0.01 (0.02)	-0.08 (0.16)	-0.04 (0.13)
<b>Smoking (Non smokers)</b>	0.05 (0.02)	0.50 (0.19)	0.52 (0.17)
<b>Status (Controls)</b>	-0.08 (0.01)	-0.87 (0.15)	-0.87 (0.13)
<b>RMSE*</b>	198.7† [145.5 ; 251.9]	193.9 [180.8 ; 207.1]	211.5 [187.0 ; 236.0]

\*: For the distribution of NETs multiplied by 1,000. Mean [min-max] over 1,000 bootstrapped samples

†: Negative predictions were censored at zero

**Table 3 : Association of rs57502213 with NETs plasma levels**

	Genotype for rs57502213		
	TC/TC	TC/-	-/-
<b>All individuals (N=657)</b>			
N	568	88	1
Mean $\pm$ SD	0.07 $\pm$ 0.13	0.15 $\pm$ 0.28	0.07
Exact zero	91 (16.0%)	13 (14.8%)	-
<b>Cases (N=372)</b>			
N	323	48	1
Mean $\pm$ SD	0.04 $\pm$ 0.06	0.12 $\pm$ 0.26	0.07
Exact zero	66 (20.4%)	9 (18.7%)	-
<b>Controls (N=285)</b>			
N	245	40	-
Mean $\pm$ SD	0.10 $\pm$ 0.18	0.20 $\pm$ 0.30	-
Exact zero	25 (10.2%)	4 (10.0%)	-

**Supplementary Table S1 : Evaluation of the control of the type I error in 10 000 bootstrap samples according to the frequency of the SNP tested with Compound Poisson-Gamma and Negative Binomial models**

	<b>Compound Poisson-Gamma</b>				<b>Negative Binomial</b>			
<b><math>\alpha</math></b>	<b>MAF 1%</b>	<b>MAF 5%</b>	<b>MAF 10%</b>	<b>MAF 20%</b>	<b>MAF 1%</b>	<b>MAF 5%</b>	<b>MAF 10%</b>	<b>MAF 20%</b>
<b>0.05</b>	0.043	0.053	0.054	0.059	0.119	0.112	0.117	0.114
<b>0.01</b>	0.007	0.010	0.013	0.013	0.039	0.036	0.038	0.038
<b>0.001</b>	0.0004	0.0014	0.0017	0.0011	0.0099	0.0077	0.0084	0.0072

**Supplementary Table S2: Evaluation of the control of the type I error in 10 000 bootstrap samples when outliers are removed, according to the frequency of the SNP tested with Compound Poisson-Gamma and Negative Binomial models**

	<b>Compound Poisson-Gamma</b>				<b>Negative Binomial</b>			
<b><math>\alpha</math></b>	<b>MAF 1%</b>	<b>MAF 5%</b>	<b>MAF 10%</b>	<b>MAF 20%</b>	<b>MAF 1%</b>	<b>MAF 5%</b>	<b>MAF 10%</b>	<b>MAF 20%</b>
<b>0.05</b>	0.053	0.054	0.053	0.052	0.059	0.048	0.046	0.043
<b>0.01</b>	0.010	0.012	0.012	0.012	0.019	0.011	0.010	0.009
<b>0.001</b>	0.0005	0.0007	0.0013	0.0017	0.0047	0.0008	0.0010	0.0014