


 Cite this: *Analyst*, 2024, **149**, 1061

 Received 28th October 2023,
Accepted 8th December 2023

DOI: 10.1039/d3an01874a

rsc.li/analyst

High-abundance peaks and peak clusters associate with pharmaceutical polymers and excipients in urinary untargeted clinical metabolomics data: exploration of their origin and possible impact on label-free quantification†

 Frank Klont,^a Fleur B. Nijdam,^a Stephan J. L. Bakker,^d
Pekka Keski-Rahkonen,^e Gérard Hopfgartner^c and TransplantLines Investigators^f

Pharmaceutical polymers and excipients represent interesting but often overlooked chemical classes in clinical exposure and bioanalytical research. These chemicals may cause hypersensitivity reactions, they can be useful to confirm exposure to pharmaceuticals, and they may pose bioanalytical challenges, including ion suppression in liquid chromatography-mass spectrometry (LC-MS)-based workflows. In this work, we assessed these chemicals in light of a rather surprising finding presented in two previously published studies, namely that usage of cyclosporine A, an immunosuppressive drug which is known to be cleared through excretion in the bile, explained the largest amount of variance in principal component analysis of urinary LC-SWATH/MS small-molecule profiling data. Specifically, we examined the freely-accessible 24-hour urine metabolomics data of 570 kidney transplant recipients included in the TransplantLines Biobank and Cohort Study (NCT03272841). These data unveiled thousands of high-abundance polymer peaks in some samples, which were associated with the use of the macrogol (*i.e.*, polyethylene glycol) 3350 oral laxative agent. In addition, we found multiple clusters of high-abundance peaks

which were linked to the exposure to two pharmaceutical excipients, namely short-chain polyethylene glycol (molecular weight <1000 Da) and polyethoxylated castor oil (also known as Kolliphor® EL or Cremophor® EL). Respectively, these excipients are used in temazepam capsules and cyclosporine A capsules, and the latter provides a plausible explanation for the rather surprising finding that instigated our work. Moreover, such explanation and our findings in general put emphasis on taking into consideration these and other pharmaceutical polymers and excipients when exploring, processing, and interpreting clinical small-molecule profiling data.

1 Introduction

Untargeted metabolomics by liquid chromatography coupled to mass spectrometry (LC-MS) is gaining momentum in clinical research, for example for identification of disease-related mechanisms, biomarkers, and druggable targets.^{1,2} This technique is furthermore used increasingly to obtain molecular evidence of chemical exposures, including dietary (*e.g.*, caffeine, ethanol), lifestyle (*e.g.*, tobacco smoke, illicit drugs), and medical exposures (*e.g.*, therapeutic drugs).^{3–6} Such evidence could enable the verification of clinical data retrieved through anamnesis and questionnaires, thus reducing various sources of bias in clinical exposure research.⁴ In addition, metabolomics can provide insights into un- and/or underreported exposures and into the real-world metabolism of xenobiotics, all representing rather underexplored application areas with considerably high clinical potential.^{5,6}

For most applications of untargeted metabolomics, researchers can benefit from both the qualitative and quantitative opportunities offered by the corresponding techniques.^{7,8} However, there are still ample challenges relating to these two opportunities preventing metabolomics from achieving its full potential in clinical research. Many of these

^aUnit of Pharmacotherapy, -Epidemiology & -Economics, Groningen Research Institute of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands. E-mail: frank.klont@rug.nl

^bDepartment of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands

^cLife Sciences Mass Spectrometry, Department of Inorganic and Analytical Chemistry, University of Geneva, Quai Ernest Ansermet 24, 1211 Geneva, Switzerland

^dDepartment of Internal Medicine, Division of Nephrology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands

^eNutrition and Metabolism Branch, International Agency for Research on Cancer (IARC/WHO), Avenue Tony Garnier 25, 69007 Lyon, France

^fGroup of Authors on Behalf of the Transplant Lines Biobank and Cohort Study, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3an01874a>



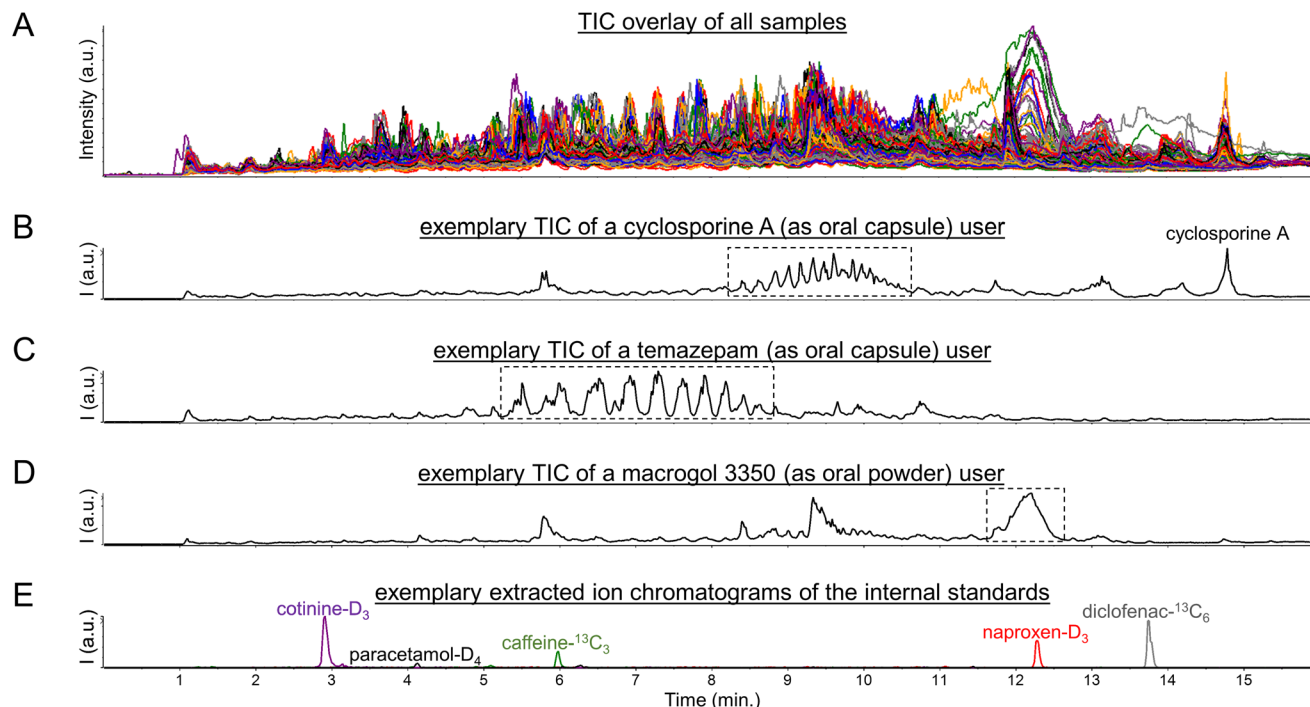


Fig. 1 Total ion current chromatograms (TIC) of (A) all 570 urine samples, (B) a user of the immunosuppressive drug cyclosporine A, (C) a user of the short-acting benzodiazepine drug temazepam, and (D) a user of the laxative agent macroglol 3350, as well as (E) MS1 extracted ion chromatograms of the internal standards which were all added at 1 pmol (0.15–0.30 ng) per μL urine corresponding to 20 pmol (3–6 ng) injected onto the LC column. The Fig. S1, S3, and S6[†] furthermore include MS1 spectra of the peaks/regions highlighted with dotted-line boxes in the Fig. 1B, C, and D.

study participants between 5 and 9 min (see Fig. 1c). The corresponding mass spectra (see Fig. S4 in the ESI[†]) did not show series of doubly-charged peaks with 7 and 22 m/z differences but rather series of singly-charged peaks with 44 m/z differences. These peaks could be identified by spectral library matching as ammonium adducts of short-chain (7- to 12-mer) polyethylene glycol molecules (see Fig. S5 in the ESI[†]). The presence of these peaks furthermore aligned reasonably well with usage of the short-acting benzodiazepine drug temazepam, as was based on the (self-reported) drug use information available in the corresponding clinical database. Such alignment seems explicable given that temazepam is often formulated as an oral capsule in which the active pharmaceutical ingredient is dispersed in short-chain polyethylene glycol (molecular weight below 1000 Da) to ensure a rapid release. This alignment also improved when using molecular evidence of (presumed) temazepam exposure, as relied on the identification of a phase II metabolite of this drug through spectral library matching (see Fig. S6 in the ESI[†]). A plausible explanation here is that temazepam is generally used on an ‘if needed’ (or: ‘*pro re nata*’, PRN) basis, while possibly also considering that it is a drug that is often misused and abused. It should be noted, however, that there is no perfect agreement between any of the two abovementioned temazepam exposure statuses and the intensity levels of signals in the peak clusters between 5 and 9 min (see Fig. S7 in the ESI[†]). Concretely, these peak clusters were not observed in some of the temaze-

pam users, which should likely be viewed in the context that temazepam can also be supplied as an oral tablet which does not contain polyethylene glycol excipients. Still, we believe that the pharmaceutical excipient is a probable cause for the series of high-abundance peaks observed between 5 and 9 min in some of the urine samples.

Besides unusual peak clusters, some samples also featured a single, large, and broad peak at 12–12.5 min (see Fig. 1d) which caught our attention when visually inspecting the TICs. The corresponding mass spectra revealed a plethora of multiply-charged peaks which roughly corresponded to monoisotopic masses in the range of 3000 to 3500 Dalton (see Fig. S8 in the ESI[†]). The latter hinted towards usage of macroglol (*i.e.*, polyethylene glycol) 3350 oral laxative agents, which was supported by the (self-reported) drug use information available in the corresponding clinical database. However, agreement between the latter information and intensity levels of signals in the peak at 12–12.5 min is far from perfect (see Fig. S9 in the ESI[†]), as may be expected given that these laxative agents are often used on a PRN basis. Still, we believe that our findings may seem to disagree with the general understanding that macroglol laxative agents are not (or only in trace amounts) absorbed from the gastrointestinal tract.¹⁶

Altogether, the abovementioned findings suggest that pharmaceutical polymers and excipients could lead to (ten) thousands of high-abundance peaks being detected in clinical metabolomics datasets. As such, these peaks could potentially



impact commonly used metabolomics data (pre)processing steps, in particular data normalization procedures, based on their high numbers and abundances but also by suppressing the signals of internal standards (see Fig. 1e). Thereby, pharmaceutical polymers and excipients may potentially affect the quantitative potential of LC-MS-based metabolomics, as will be discussed below.

3.2 Selection of exemplary quantifier signals

To assess the quantitative performance of the workflow that was used to obtain the data studied in this work, we selected two target analytes, namely the endogenous muscle breakdown product creatinine and the exogenous compound cotinine which is a phase I metabolite of nicotine. These compounds were the only two compounds which could both be detected in the metabolomics datasets and for which quantitative data was already available in the clinical database through prior performance of routine measurements in ISO 15189 certified laboratories.

Next, we evaluated the correlations between (unnormalized) signals in the metabolomics dataset and the available 'reference' data (see Fig. 2). These evaluations yielded fairly linear correlations for the MS1-level precursor ions while various degrees of saturation were observed for the MS2-level residual precursor ions and representative MS2-level fragment ions. A possible explanation for this phenomenon is the use of the so-called 'ion transmission control' feature which could only be enabled for MS1-level acquisition on the MS instrument used to obtain the metabolomics data. Consequently, using the MS1-level precursor ions seems preferable for quantitative applications, although the linearity of certain lower-abundance MS2-level signals (*e.g.*, fragment ions, isotope peaks) demonstrated better linearity as compared to the MS2-level residual precursor ions (see Fig. S10 in the ESI†). Besides being useful for qualitative purposes, all MS2-level signals can furthermore be useful to assess specificity of the MS1-level precursor data, as can contribute to verifying data reliability.

At last, we selected creatinine and cotinine based on the presence of reference data, yet it should be acknowledged that reasons for disagreement between data from different assays come in plenty. For cotinine, we could yield some insights into such (potential) disagreement given that a deuterated version of cotinine was added to all samples as internal standard. This addition allowed us to correct for various sources of analytical variability, and the corresponding light-to-heavy (relative) feature ratios showed a linear and markedly stronger correlation with the reference data as compared to the absolute feature area (see Fig. 3). The latter data can, however, still be very useful for quantitative purposes, yet sufficient attention should be paid to statistical power in light of the magnitude of differences that can meaningfully be detected between study groups.¹⁷

3.3 Effect of normalization on exemplary quantifier signals

The unnormalized MS1-level data for creatinine and cotinine were scaled based on several normalization approaches, namely most likely ratio-based normalization ("MLR"),¹⁴ total

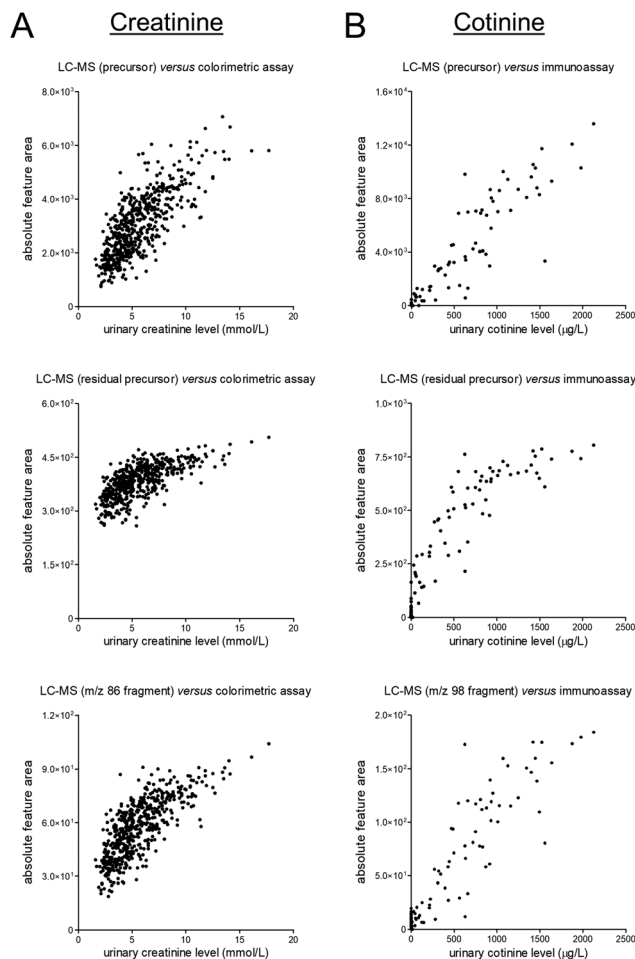


Fig. 2 Scatter plots presenting unnormalized feature data of the MS1-level precursors (top), MS2-level residual precursors (middle), and representative MS2-level fragments (bottom) on the y-axis and the results of previously conducted routine measurements (in an ISO 15189 certified laboratory) on the x-axis for (A) the endogenous muscle breakdown product creatinine and (B) the exogenous phase I nicotine metabolite cotinine. The corresponding linear regression summary statistics for the MS1-level data are presented in Table S2.†

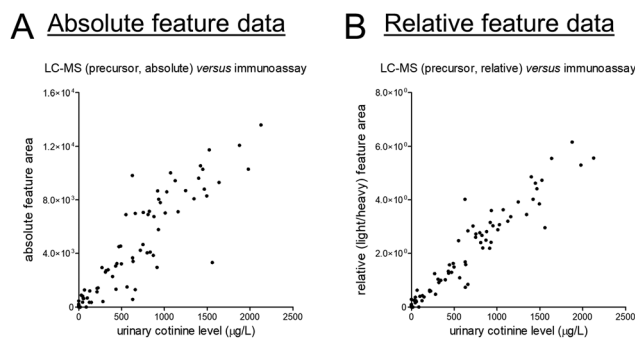


Fig. 3 Scatter plots presenting unnormalized feature data of the (A) MS1-level precursor and (B) the light/heavy ratio of the MS1-level precursors on the y-axis and the results of previously conducted routine measurements (in an ISO 15189 certified laboratory) on the x-axis for the exogenous phase I nicotine metabolite cotinine. The corresponding linear regression summary statistics are presented in Table S2.†



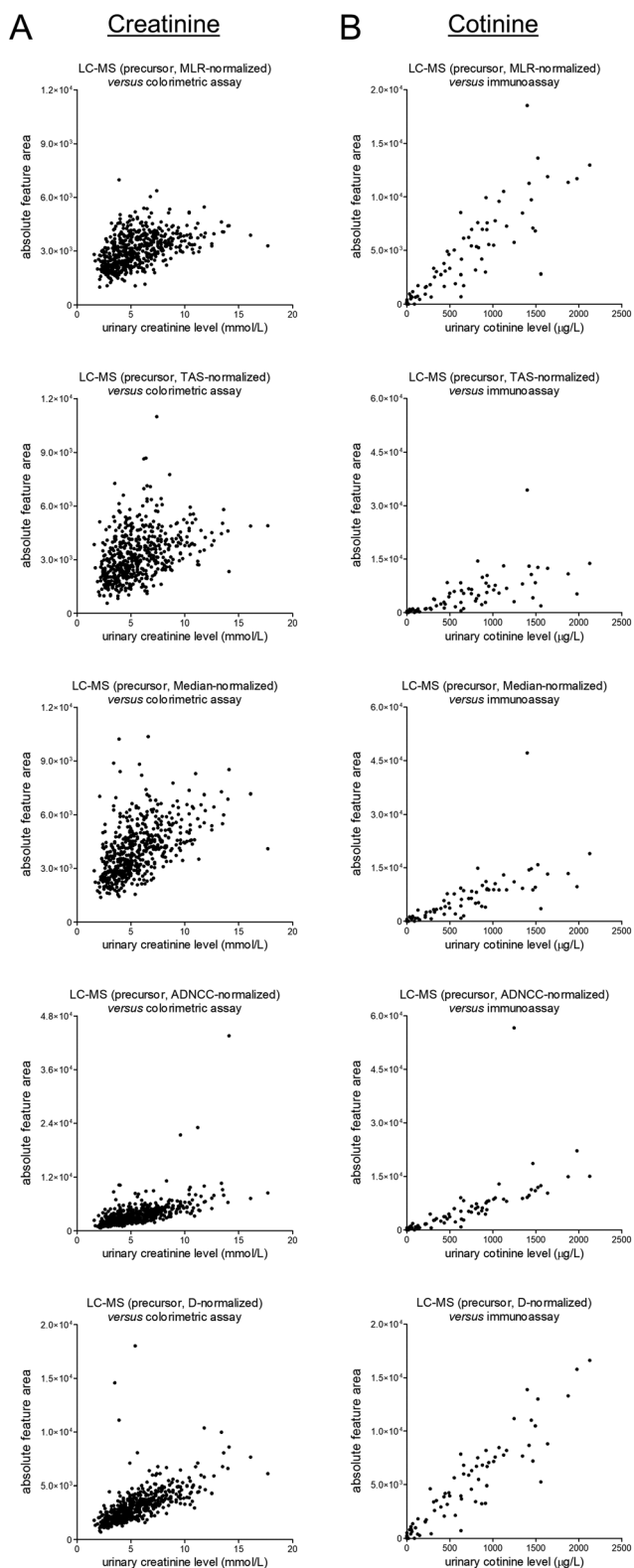


Fig. 4 Scatter plots presenting differentially-normalized feature data of the MS1-level precursors (from top to bottom: MLR, TAS, Median, ADNCC, and D) on the y-axis and the results of previously-conducted routine measurements (in an ISO 15189 certified laboratory) on the x-axis for (A) the endogenous muscle breakdown product creatinine and (B) the exogenous phase I nicotine metabolite cotinine. The corresponding linear regression summary statistics are presented in Table S2.†

area sums-based normalization (“TAS”), median peak ratios-based normalization (“Median”), normalization based on all five included internal standards (“ADNCC”), and normalization based on the diclofenac- $^{13}\text{C}_6$ internal standard only (“D”). The normalized data were subsequently plotted against the earlier-mentioned reference data, and the corresponding scatter plots are shown in Fig. 4. Upon comparison with matching plots for the unnormalized data, which were already shown in Fig. 2, all of the normalization procedures led to worsening of the correlations. All procedures furthermore led to (apparently) outlying results, although this effect seems less pronounced for MLR-based normalization.

Next, we aimed to assess whether the high-abundance peak clusters associated with the (presumed) exposure to CsA, temazepam, and macrogol 3350 contributed to the abovementioned worsening of the correlations, for example through the large numbers and high intensity levels of the respective peaks or by potentially suppressing the signals of internal standards. For this purpose, we provided views of the plots presented in Fig. 4 in which coloring is applied to indicate differential exposure statuses based on identified cyclosporine A, identified temazepam, and self-reported use of macrogol 3350 (see Fig. S11 and S22 in the ESI†). In our view, none of these plots seems to indicate a noticeable contribution of these exposures to the apparent worsening of the correlations when scaling the data, with the exception of ‘ADNCC’ multi-internal standard-based approach. In the corresponding figures (see Fig. S15 and S21 in the ESI†), one might notice some of the macrogol 3350-positive samples being outliers which warrants further investigation.

At last, in spite of not noticing a worrying effect of exposure to CsA, temazepam, and macrogol 3350 on data normalization, we would still feel inclined not to apply any normalization procedure to the studied clinical dataset, at least none of the studied approaches, given the findings depicted in Fig. 4. We are aware of other and more sophisticated normalization techniques, and we furthermore understand the necessity to normalize when pooling metabolomics data across studies.^{18,19} In addition, we acknowledge that typical metabolomics data processing pipelines include a data filtering step, often using the ‘80% rule’ which removes features that have missing data in more than 20% of the samples,²⁰ before performing normalization. However, the use of normalization factors obtained after such a data filtering procedure did not seem to affect the observed correlations that much, at least not for this dataset (see Fig. S23 in the ESI†).

4 Conclusions

This work provides a possible explanation for the unexpected observation put forward by two previously published studies that usage of cyclosporine A, an immunosuppressive drug which should predominantly be excreted in the bile, was responsible for the most pronounced clustering in principal component analysis of urinary metabolomics data from kidney transplant recipients. Specifically, we found series of high-abundance peaks with 14 and 44 Dalton differences which



likely originate from the lipid-based pharmaceutical excipient used to enhance oral delivery of this drug. Comparable series were also found in the urine of temazepam users, and the corresponding signals likely originate from the polymeric pharmaceutical ingredient polyethylene glycol that is used to ensure a rapid release of this short-acting benzodiazepine drug. In addition, we found clusters of thousands of high-abundance peaks corresponding to monoisotopic masses in the range of 3000 to 3500 Dalton. These findings hinted towards usage of macrogol 3350 oral laxative agents, as was supported by available (self-reported) drug use information. However, it should be noted that such an explanation goes against the consensus that these agents are not (or only in trace amounts) absorbed from the gastrointestinal tract and thus are not expected in urine.

Besides describing the potential presence of pharmaceutical polymers and excipients in clinical metabolomics data, our work also aimed to assess the potential implications of their presence on the quantitative performance of LC-MS-based small-molecule profiling workflows. In particular, we put emphasis on the selected data normalization techniques which did not lead to more reliable quantitative data, at least not in our opinion and not for the dataset we studied. Data normalization is, however, an important step in metabolomics data processing pipelines, and we believe that our findings can help to optimize existing and/or develop novel data normalization approaches to be applied to clinical metabolomics datasets.

Author contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. Conceptualization – F. K., G. H.; methodology – F. K.; formal analysis – F. K.; investigation – F. K., F. B. N., P. K-R.; resources – S. J. L. B., T. I., G. H.; data curation – F. K.; writing – original draft – F. K.; writing – review & editing – F. K., F. B. N., S. J. L. B., P. K-R., G. H.; visualization – F. K.; supervision – F. K., S. J. L. B., T. I., P. K-R., G. H.; funding acquisition – F. K., S. J. L. B., T. I., G. H.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Frank Klont received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Actions grant agreement No. 887661 and from the Netherlands Organisation for Scientific Research NWO (domain Applied and Engineering Sciences) under the Veni grant agreement No. 19060. The funders had no role in the study design, data collection, analysis, reporting, or the decision to submit for publication.

The TransplantLines Biobank and Cohort study was supported by grants from Astellas BV (project code: TransplantLines Biobank and Cohort study), Chiesi Pharmaceuticals BV (project code: PA-SP/PRJ-2020-9136), and NWO/TTW *via* a partnership program with DSM, Animal Nutrition and Health, The Netherlands (project code: 14939). The project was co-financed by the Dutch Ministry of Economic Affairs and Climate Policy by means of so-called PPP-allowances, made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships (project code: PPP-2019-032 and PPP-2022-015). The funders had no role in the study design, data collection, analysis, reporting, or the decision to submit for publication.

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

TransplantLines Investigators: Coby Annema, Stephan J. L. Bakker, Stefan P. Berger, Hans Blokzijl, Frank A. J. A. Bodewes, Marieke T. de Boer, Kevin Damman, Martin H. de Borst, Arjan Diepstra, Gerard Dijkstra, Caecilia S. E. Doorenbos, Rianne M. Douwes, Michele F. Eisenga, Michele E. Erasmus, C. Tji Gan, Antonio W. Gomes Neto, Eelko Hak, Bouke G. Hepkema, Frank Klont, Tim J. Knobbe, Daan Kremer, Henri G. D. Leuvenink, Willem S. Lexmond, Vincent E. de Meijer, Hubert G. M. Niesters, Gertrude J. Nieuwenhuijs-Moeke, L. Joost van Pelt, Robert A. Pol, Adelita V. Ranchor, Jan Stephan F. Sanders, Marion J. Siebelink, Riemer J. H. J. A. Slart, J. Casper Swarte, Daan J. Touw, Marius C. van den Heuvel, Coretta van Leer-Buter, Marco van Londen, Erik A. M. Verschuuren, Michel J. Vos and Rinse K. Weersma.

References

- 1 A. Palermo, *Drug Discovery Today*, 2023, **28**, 103460.
- 2 N. P. Long, T. D. Nghi, Y. P. Kang, N. H. Anh, H. M. Kim, S. K. Park and S. W. Kwon, *Metabolites*, 2020, **10**, 51.
- 3 F. Klont, S. Jahn, C. Grivet, S. König, R. Bonner and G. Hopfgartner, *Talanta*, 2020, **211**, 120747.
- 4 F. Klont, D. Kremer, A. W. Gomes Neto, S. P. Berger, D. J. Touw, E. Hak, R. Bonner, S. J. L. Bakker and G. Hopfgartner, *J. Clin. Epidemiol.*, 2021, **135**, 10–16.
- 5 F. Klont, S. Stepanović, D. Kremer, R. Bonner, D. J. Touw, E. Hak, S. J. L. Bakker and G. Hopfgartner, *Food Chem. Toxicol.*, 2022, **165**, 113188.
- 6 F. Klont, P. Sosnowski, D. Kremer, T. J. Knobbe, R. Bonner, H. Blokzijl, R. K. Weersma, S. J. L. Bakker, T. Investigators, E. Hak, D. J. Touw and G. Hopfgartner, *Metabolites*, 2022, **12**, 942.
- 7 R. Bonner and G. Hopfgartner, *TrAC, Trends Anal. Chem.*, 2019, **120**, 115278.
- 8 M. Raetz, R. Bonner and G. Hopfgartner, *Metabolomics*, 2020, **16**, 71.



- 9 B. B. Misra, *Eur. J. Mass Spectrom.*, 2020, **26**, 165–174.
- 10 X. Du, F. Dastmalchi, H. Ye, T. J. Garrett, M. A. Diller, M. Liu, W. R. Hogan, M. Brochhausen and D. J. Lemas, *Metabolomics*, 2023, **19**, 11.
- 11 A. Verhoeven, M. Giera and O. A. Mayboroda, *Analyst*, 2020, **145**, 3801–3808.
- 12 G. J. Burckart, T. E. Starzl, R. Venkataramanan, H. Hashim, L. Wong, P. Wang, L. Makowka, A. Zeevi, R. J. Ptachcinski and J. E. Knapp, *Transplant Proc.*, 1986, **18**, 46–49.
- 13 M. F. Eisenga, A. W. Gomes-Neto, M. van Londen, A. L. Ziengs, R. M. Douwes, S. P. Stam, M. C. J. Osté, T. J. Knobbe, N. R. Hessels, A. M. Buunk, C. Annema, M. J. Siebelink, E. Racz, J. M. Spikman, F. A. J. A. Bodewes, R. A. Pol, S. P. Berger, G. Drost, R. J. Porte, H. G. D. Leuvenink, K. Damman, E. A. M. Verschuuren, V. E. de Meijer, H. Blokzijl and S. J. L. Bakker, *BMJ Open*, 2018, **8**, e024502.
- 14 J. P. Lambert, G. Ivosev, A. L. Couzens, B. Larsen, M. Taipale, Z. Y. Lin, Q. Zhong, S. Lindquist, M. Vidal, R. Aebbersold, T. Pawson, R. Bonner, S. Tate and A. C. Gingras, *Nat. Methods*, 2013, **10**, 1239–1245.
- 15 J. S. Abraham, F. R. Bentley, R. N. Garrison and H. M. Cryer, *Transplantation*, 1991, **52**, 101–105.
- 16 F. Savino, S. Viola, M. Erasmo, G. Di Nardo, S. Oliva and S. Cucchiara, *BMC Pediatr.*, 2012, **12**, 178.
- 17 B. J. Blaise, G. Correia, A. Tin, J. H. Young, A. C. Vergnaud, M. Lewis, J. T. M. Pearce, P. Elliott, J. K. Nicholson, E. Holmes and T. M. D. Ebbels, *Anal. Chem.*, 2016, **88**, 5179–5188.
- 18 P. Cuevas-Delgado, D. Dudzik, V. Miguel, S. Lamas and C. Barbas, *Anal. Bioanal. Chem.*, 2020, **412**, 6391–6405.
- 19 V. Viallon, M. His, S. Rinaldi, M. Breeur, A. Gicquiau, B. Hemon, K. Overvad, A. Tjønneland, A. L. Rostgaard-Hansen, J. A. Rothwell, L. Lecuyer, G. Severi, R. Kaaks, T. Johnson, M. B. Schulze, D. Palli, C. Agnoli, S. Panico, R. Tumino, F. Ricceri, W. M. M. Verschuren, P. Engelfriet, C. Onland-Moret, R. Vermeulen, T. H. Nøst, I. Urbarova, R. Zamora-Ros, M. Rodriguez-Barranco, P. Amiano, J. M. Huerta, E. Ardanaz, O. Melander, F. Ottoson, L. Vidman, M. Rentoft, J. A. Schmidt, R. C. Travis, E. Weiderpass, M. Johansson, L. Dossus, M. Jenab, M. J. Gunter, J. Lorenzo Bermejo, D. Scherer, R. M. Salek, P. Keski-Rahkonen and P. Ferrari, *Metabolites*, 2021, **11**, 631.
- 20 S. Bijlsma, I. Bobeldijk, E. R. Verheij, R. Ramaker, S. Kochhar, I. A. Macdonald, B. Van Ommen and A. K. Smilde, *Anal. Chem.*, 2006, **78**, 567–574.

