Supplementary Text

Profiles of histidine-rich glycoprotein associate with age and risk of all-cause mortality

*Mun-Gwan Hong, Tea Dodig-Crnković, Xu Chen, Kimi Drobin, Woojoo Lee, Yunzhang Wang, Fredrik Edfors, David Kotol, Cecilia E. Thomas, Ronald Sjöberg, Jacob Odeberg, Anders Hamsten, Angela Silveira, Per Hall, Peter Nilsson, Yudi Pawitan, Mathias Uhlén, Nancy L. Pedersen, Sara Hägg, Patrik K. E. Magnusson and Jochen M. Schwenk*

**Table of content**

[1. Samples and selections 3](#_Toc44074365)

[a) Sample set 6 from IMPROVE cohort 3](#_Toc44074366)

[b) Sample set 7 from SCARF study 3](#_Toc44074367)

[c) Sample set 8 from CHAPS 3](#_Toc44074368)

[d) Sample set 9 from Karma study 4](#_Toc44074369)

[e) Sample selection for replication 4](#_Toc44074370)

[2. Suspension Bead Array Assays 4](#_Toc44074371)

[a) Antibody selection 4](#_Toc44074372)

[b) Suspension Bead Array Procedure 5](#_Toc44074373)

[c) Assay design 5](#_Toc44074374)

[d) SBA data quality control and preprocessing 6](#_Toc44074375)

[e) Data acquisition of replication sample sets 6](#_Toc44074376)

[3. Sandwich immunoassays 6](#_Toc44074377)

[4. Protein and peptide microarray analysis 7](#_Toc44074378)

[5. Mass spectrometry analysis 7](#_Toc44074379)

[6. References 9](#_Toc44074380)

## Samples and selections

### Sample set 6 from IMPROVE cohort

IMPROVE is a multicenter, longitudinal, observational cohort study of individuals at high-risk of cardiovascular diseases, from seven recruiting centers in five European countries: Finland, Sweden, the Netherlands, France and Italy (Baldassarre et al, 2010). A total of 3,711 participants were enrolled between March 2004 and April 2005. Eligibility criteria included age from 55 to 79 years, presence of at least three cardiovascular risk factors and absence of symptoms of cardiovascular diseases. Cases denote individuals who suffered a vascular event from month 15 to month 36 during follow-up, while controls did not have any vascular event during the time. Forty-three cases and age-, sex- and center-matched controls were used in the present study.

### Sample set 7 from SCARF study

The Stockholm Coronary Atherosclerosis Risk Factor (SCARF) study included survivors of a first myocardial infarction (MI) before the age of 60 years and age- and sex-matched control subjects free from the disease at enrollment, between January 1996 and December 2000 (Samnegård et al, 2005). Cases were patients admitted to the three hospitals in the northern part of Stockholm area with an ongoing cardiac event, and controls were recruited in parallel from the general population of the same residence area. The blood samples were collected 3 months after the index cardiac event.

### Sample set 8 from CHAPS

In the Carlscrona Heart Attack Prognosis Study (CHAPS), 5292 human subjects were consecutively recruited and EDTA plasma samples were collected in a coronary intensive care unit during 1992 to 1996 (Odeberg et al, 2014). Of these 908 patients aged 30–74 years had at discharge received the diagnosis of either myocardial infarction (527) or unstable angina (381). As control group, 948 patients aged 30–74 years were identified who were admitted with suspected acute coronary syndrome (ACS), but were subsequently diagnosed as non-ACS and, furthermore, were not diagnosed with stable CAD. From this group, forty-three ACS patients and an equal number of non-ACS controls were selected for a nested case-control discovery study. Those controls were analyzed in the present study. The activated partial thromboplastin time (aPTT) was analysed in Sodium Citrate blood samples using a routine diagnostic method on a Trombotrack instrument (Nycomed, Norway). The procedures for blood sampling and laboratory analyses followed the routines of the Department of Clinical Chemistry at Blekinge County Hospital and analyses were performed in the certified hospital laboratory using fresh samples collected at hospital admission. The hospital laboratory reference range with this assay was APTT 22-32 s.

### Sample set 9 from Karma study

The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA; www.karmastudy.org) is a prospective population-based cohort (Gabrielson et al, 2017). Nearly 71,000 women were included in the cohort between 2011-2013. Women invited for mammography screening or clinical mammography at any of the four mammography units in Sweden were recruited into the cohort study. At study entry, information on a vast number of lifestyle factors was gathered through a web-based questionnaire. Plasma samples from non-diseased women were selected from mammographic breast density measurements (absolute volumetric density, cm3), where 300 dense and 300 non-dense samples were matched on age (median 54, 40-73) and body mass index (median 24, 19-30).

### Sample selection for replication

The subjects of the sets 2 to 9 comprised of 829 subjects from non-diseased control groups and 3,203 chosen to reflect Swedish population. The entire set of subjects were from 9 to 93 years old at blood draw. Blood samples had been prepared either as serum or plasma (Table 1, supplementary Fig S2). The sample set 2 (204 subjects) was chosen from a population-based prospective cohort in Sweden (Almqvist et al, 2011) and the set 3 (2999 subjects) was from the twin cohort same as the set 1 (Lichtenstein et al, 2002; Magnusson et al, 2013), in which disease status was not considered during recruitment. Two other sample sets (set 4 and 5) included 100 subjects that were selected from cancer-related studies and derived from the same twin cohort as the set 1 (Lichtenstein et al, 2002; Magnusson et al, 2013). Forty-four subjects in the set 3 and one in the set 4 were overlapped with the set 1. The data of those 45 subjects were excluded in the meta-analysis. Sample sets 6 to 9 (729 subjects) were from four independent studies (Table 1) (Baldassarre et al, 2010; Gabrielson et al, 2017; Odeberg et al, 2014; Samnegård et al, 2005).

## Suspension Bead Array Assays

### Antibody selection

Antibodies from the Human Protein Atlas (HPA) with a concentration above 0.05 mg/ml and passing specificity assessment on planar protein microarrays (Sjoberg et al, 2012) were used for the analysis of sample sets 1 and 2. No other protein related criteria other than antibody availability at the time of the study were applied. For the analysis of sample sets 4, 5, 6, 7, and 8, antibodies were selected without considering target proteins. The antibodies for sample set 3 included antibodies that had been used and showed some potentially interesting results in other studies for various diseases. Two of them were HPA045005 from the HPA and BSI0137 (Batch ID 0137090310), which is a monoclonal mouse antibody from BioSystems International Kft (Guergova-Kuras et al, 2011). The antibodies for sample set 9 were chosen for a mammography study for its own interest (Bystrom et al, 2018).

### Suspension Bead Array Procedure

Coupling of antibodies to magnetic beads (MagPlex, Luminex Corp) was performed as previously described (Byström et al, 2014) using 1.6 µg of a given antibody per 500 000 beads. For this purpose, antibodies were diluted using a liquid handler (EVO150, TECAN), beads were washed on a magnet using a plate washer (EL406, Biotek). Coupled beads were blocked and stored in buffer (Blocking reagent for ELISA, Roche) supplemented with ProClin (Sigma) at 4°C in the dark. Equal volumes of beads carrying different capture antibodies were mixed to create a suspension bead array (SBA). As described before, samples were labeled with biotin and heat-treated (Schwenk et al, 2010). In short, samples were centrifuged, diluted 1/10 in PBS using a liquid handler (SELMA, CyBio AG), labeled with 10 mg/ml NHS-biotin (Pierce), which was quenched using 0.5 M Tris-HCl. Labeled samples were diluted 1:50 in assay buffer (0.5% (w/v) polyvinyl alcohol and 0.8% (w/v) polyvinylpyrrolidone (Sigma) in 0.1% casein in PBS supplemented with 0.5 mg/ml rabbit IgG (Bethyl Laboratories) and heat-treated for 30 min at 56°C using thermocyclers. Then 45 µl of samples were added to 5 µl SBA, incubated overnight at room temperature with constant rotation at 650 rpm (Grant). Thereafter, beads were washed 3x in 100 µl PBS-T and 50 µl 0.4% PFA was added for 10 min. Beads were washed once again with PBS-T before streptavidin R-PE (Invitrogen) was diluted 1:750 in PBS-T and 50 µl was added for 20 min. Prior to analysis in FlexMap3D instrument (Luminex Corp), beads were washed 3x in 100 µl PBS-T. Median signal intensities (MFI) of each bead ID were used for subsequent data analysis.

### Assay design

All 372 samples from sample sets 1 and 2 were together randomly allocated into wells in four 96-well plates. One sample from sample set 1 and one from sample set 2 were loaded into two more wells as a repeated control within a plate. Another sample in each cohort was transferred to two more wells of two different plates as a control to examine inter-plate variation. The data of each of those 4 samples was combined by taking mean of three measures. All the human materials were biotinylated together with four negative controls that contained only buffer. For the entire 19 assays for discovery stage, the samples were labeled two times.

The selected antibodies were divided into collections of 384 antibodies including positive and negative controls, anti-albumin and no antibody, respectively. These antibodies were then coupled onto beads and used to create a suspension bead array (SBA). For discovery, the selection of the affinity binders for one SBA was determined by technical reasons such as the available amount. Every antibody in an SBA was coupled with beads with a different color code as detailed together with the assay procedure in the Supplementary Information and as described earlier (Byström et al, 2014). This assay provided protein profiles from up to 384 antibodies and 384 samples per batch.

### SBA data quality control and preprocessing

Because an aliquot of mixed bead solution was suspended into each sample, all values of the samples that were seemingly failed within an assay (with 384 antibodies) were discarded rather than one measure of a sample for an antibody. These were samples 1) that had median bead counts lower than twenty, 2) that had median values of MFIs lower than the median of the negative controls (buffer only) in the same plate and assay, and 3) that were detected as an outlier by robust PCA using ‘rrcov’ R package (version 1.3-4)(Hubert et al, 2005). The cutoff probability values in an outlier diagnostic plot were 0.025 for both score and orthogonal distance coordinates. Those deviating beyond the cutoffs in both coordinates were classified as outliers, setting alpha, the proportional tolerance, to 0.9.

The human samples were of two different types in terms of preparation method, plasma and serum. The two blood preparation types showed considerable dissimilarity, which was expected (supplementary Fig S2). Since such contrast was not of our research interest here, the data was split by the sample preparation type after quality control, when two types of samples were analyzed in the same assay plates.

The variation across sample plates was minimized by Multi-MA method, with the assumption that the mean of observed values for each antibody within a plate is same as those of the other plates (Hong et al, 2016). The means of log-transformed measures within each plate were positioned in a 4-dimensional space, in which each axis corresponded to one plate. The vector that goes through origin and (1, 1, 1, 1) is named A. The projection of each point onto the A axis is computed. All values for the point were shifted as much as the element vector on the corresponding plate axes of the projection. The proteomic data for pQTL analysis was pre-processed including technical controls and using probabilistic quotient normalization (PQN) (Kato et al, 2011) before the Multi-MA normalization.

### Data acquisition of replication sample sets

Data of other replication samples (sample sets 3-9) were acquired using the same protocol with a few variations. For each original study for sample sets 4-8, the samples were distributed into plates together with patient samples. The 383 other antibodies selected for each of the intended studies were included in the assays. Experiment and data preprocessing were conducted together with those additional samples and antibodies. Data of disease-free controls and for HPA045005 were extracted from the processed full data sets.

## Sandwich immunoassays

To confirm the capture of HRG by HPA045005, a sandwich immunoassay was developed and used for the detection of a full-length recombinant HRG protein, which was a kind gift from Hanna Tegel and Johan Rockberg (AlbaNova University Center, KTH). A multiplexed SBA containing a library of HPA antibodies including HPA045005 and the anti-HRG antibody HPA054598 were used as capture reagents, as described previously (Häussler et al, 2019). An empty bead and one rabbit IgG bead were also included in the SBA for measuring background signal. For detection, the antibody HPA054598 was biotinylated as previously described (Dezfouli et al, 2014). The HRG protein was prepared in a four-fold serial dilution (500 ng/ml to 0.1 ng/ml) in PVXC buffer (0.1% casein, 0.5% (w/v) polyvinylalcohol, 0.8% (w/v) polyvinylpyrrolidone, Sigma- Aldrich) supplemented with 0.5 mg/ml purified rabbit IgG (Bethyl laboratories). The serial dilution was added in triplicate to a 96-well plate along with three control PVXC buffer wells. Five microliters of the SBA were incubated overnight with 45 µl protein standard. Following that, the beads were washed three times in PBS-T 0.05%. The biotinylated detection antibody was diluted to 1 µg/ml in PBS-T 0.05% and 25 µl was transferred to the beads (SELMA, Cybio). After 1.5 h incubation at room temperature, the beads were washed three times in PBS-T 0.05%. R-phycoerythring-labeled streptavidin (Invitrogen) was diluted 1:500 in PBS-T 0.05% and 50 µl incubated with the SBA for 20 min. Finally, beads were washed three times and measured in PBS-T with a FlexMap 3D instrument. Presence of detection antibody was confirmed with an anti-rabbit IgG bead.

## Protein and peptide microarray analysis

To determine the selectivity of the antibody binding, a set of 16,728 protein fragments from the Human Protein Atlas were used to generated microarrays with 21,120 features corresponding representing 12,412 unique Ensemble Gene IDs according to a previous protocol (Sjoberg et al, 2012). The array contained the protein fragments of a length between approximately 20-150 residues and included the antigen used to generate HPA045005

We also aimed at determining the binding of HPA045005 to peptides representing its antigen using high density peptide arrays from NimbleGen using a previously described protocol. (Forsstrom et al, 2014). The array a set of 12-mer peptides with 11 residue overlap representing the antigen did not reveal a significantly prominent recognition of these peptides above background.

## Mass spectrometry analysis

An LC-MS/MS proteomics experiment was carried out to verify the findings from the affinity proteomics experiments orthogonally. A serum sample pool from 3 males and 2 females was diluted 5 times with 1x PBS. Sodium deoxycholate (SDC) was added to a final concentration of 1% (w/v) and proteins reduced and alkylated for 10 min at 96°C in 10 mM dithiothreitol and for 30 min at room temperature in the dark in 50 mM chloroacetamide for 30 min. The sample was diluted with 1x PBS to the final SDC concentration of 0.1% (w/v) and split into 5 tubes so each tube contained 1 uL of raw plasma. Protein digestion was performed using 5 different enzymes (supplementary Table S4) to increase the peptide coverage by data dependent acquisition (Tsiatsiani &Heck, 2015). Formic acid was added to quench the enzymatic reaction to the final concentration of 1% (v/v) and SDC was let to precipitate for 30 min at room temperature. Samples were centrifuged at 16 400RCF for 5 min and subjected to LC-MS/MS analysis.

Samples were analyzed using the UltiMate 3000 capillary-liquid chromatography system (Thermo Scientific) with an EASY‐Spray ion source connected to Q Exactive HF (Thermo Scientific) mass spectrometer. In total, 3.5 µg of peptides were loaded onto PepMap100 trap column (300 μm × 5 mm, C18, 5μm, 100 Å, Thermo Scientific), washed 5 min at 15 μl/min with 100% of Solvent A (3% ACN, 97% H20, 0.1% FA) and separated with PepMap RSLC C18 (150 µm x 15 cm, 2 µm, 100 Å, Thermo Scientific) analytical column using linear 60 min gradient of 1-32% Solvent B (95% ACN, 5% H20, 0.1% FA) at a flow rate of 3.6 µl/min. The analytical and trap columns were kept at 55°C by the in-source temperature controller and 40°C by the column oven temperature controller respectively.The MS analysis was performed using a Top10 method starting with an MS1 scan performed at resolution of 60,000 (mass range 350–1,200 m/z, AGC 3e6, max IT 100 ms) and followed by ten consecutive MS2 scans at resolution of 30,000 (AGC 2e5, max IT 105 ms) with normalized collision energy set to 28.

Resulting raw files were searched using MaxQuant version 1.6.1.0 (Cox &Mann, 2008) with the built-in search engine Andromeda against wild type sequence of HRG and HRG sequence that contained the four SNP variants (rs10770, rs9898, rs2228243, rs1042464). Search parameters were set to peptide length ranging 7-25 amino acids, enzyme specificity was according to the used enzyme for digestion (supplementary Table S4) with maximum of two mis-cleavages, 4.5 ppm match tolerance for precursor ions and 20 ppm for fragment ions with 1% false discovery rate on both the peptide and protein level. Carbamidomethylation on cysteine was set as a fixed modification.

The assay was able to detect one region that included the N493I dbSNP (rs1042464). Additionally, data from the Peptide Atlas was used to look for evidence of the missing SNPs. It confirmed both that the rs1042464 variant has been observed and yield useful proteotypic peptides suitable for LC-MS/MS analysis as well as the three other SNPs are never found (rs10770, rs2228243) or very unlikely to be observed (rs9898) (supplementary Fig S8). This is probably due to the properties of peptides being unsuitable for LC-MS/MS analysis in terms of poor ionization or due to the matrix effects of plasma proteome.

## References

Almqvist, C., Adami, H.-O., Franks, P.W., Groop, L., Ingelsson, E., Kere, J., Lissner, L., Litton, J.-E., Maeurer, M., Michaëlsson, K.*, et al.* (2011). LifeGene--a large prospective population-based study of global relevance. *Eur J Epidemiol* 26: 67-77.

Baldassarre, D., Nyyssönen, K., Rauramaa, R., de Faire, U., Hamsten, A., Smit, A.J., Mannarino, E., Humphries, S.E., Giral, P., Grossi, E.*, et al.* (2010). Cross-sectional analysis of baseline data to identify the major determinants of carotid intima-media thickness in a European population: the IMPROVE study. *Eur Heart J* 31: 614-622.

Byström, S., Ayoglu, B., Häggmark, A., Mitsios, N., Hong, M.G., Drobin, K., Forsström, B., Fredolini, C., Khademi, M., Amor, S.*, et al.* (2014). Affinity proteomic profiling of plasma, cerebrospinal fluid, and brain tissue within multiple sclerosis. *J Proteome Res* 13: 4607-4619.

Bystrom, S., Eklund, M., Hong, M.G., Fredolini, C., Eriksson, M., Czene, K., Hall, P., Schwenk, J.M., and Gabrielson, M. (2018). Affinity proteomic profiling of plasma for proteins associated to area-based mammographic breast density. *Breast Cancer Res* 20: 14.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-1372.

Dezfouli, M., Vickovic, S., Iglesias, M.J., Nilsson, P., Schwenk, J.M., and Ahmadian, A. (2014). Magnetic bead assisted labeling of antibodies at nanogram scale. *Proteomics* 14: 14-18.

Forsstrom, B., Axnas, B.B., Stengele, K.P., Buhler, J., Albert, T.J., Richmond, T.A., Hu, F.J., Nilsson, P., Hudson, E.P., Rockberg, J.*, et al.* (2014). Proteome-wide epitope mapping of antibodies using ultra-dense peptide arrays. *Mol Cell Proteomics* 13: 1585-1597.

Gabrielson, M., Eriksson, M., Hammarstrom, M., Borgquist, S., Leifland, K., Czene, K., and Hall, P. (2017). Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). *Int J Epidemiol* 46: 1740-1741g.

Guergova-Kuras, M., Kurucz, I., Hempel, W., Tardieu, N., Kadas, J., Malderez-Bloes, C., Jullien, A., Kieffer, Y., Hincapie, M., Guttman, A.*, et al.* (2011). Discovery of lung cancer biomarkers by profiling the plasma proteome with monoclonal antibody libraries. *Mol Cell Proteomics* 10: M111 010298.

Häussler, R.S., Bendes, A., Iglesias, M., Sanchez-Rivera, L., Dodig-Crnković, T., Byström, S., Fredolini, C., Birgersson, E., Dale, M., Edfors, F.*, et al.* (2019). Systematic Development of Sandwich Immunoassays for the Plasma Secretome. *Proteomics* 19: e1900008.

Hong, M.-G., Lee, W., Nilsson, P., Pawitan, Y., and Schwenk, J.M. (2016). Multidimensional Normalization to Minimize Plate Effects of Suspension Bead Array Data. *J Proteome Res* 15: 3473-3480.

Hubert, M., Rousseeuw, P.J., and Branden, K.V. (2005). ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics* 47: 64-79.

Kato, B.S., Nicholson, G., Neiman, M., Rantalainen, M., Holmes, C.C., Barrett, A., Uhlen, M., Nilsson, P., Spector, T.D., and Schwenk, J.M. (2011). Variance decomposition of protein profiles from antibody arrays using a longitudinal twin model. *Proteome Science* 9.

Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P., and Pedersen, N.L. (2002). The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J Intern Med* 252: 184-205.

Magnusson, P.K.E., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H., Halldner, L., Lundström, S., Ullén, F., Långström, N.*, et al.* (2013). The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res Hum Genet* 16: 317-329.

Odeberg, J., Freitag, M., Forssell, H., Vaara, I., Persson, M.L., Odeberg, H., Halling, A., Rastam, L., and Lindblad, U. (2014). The influence of smoking and impaired glucose homoeostasis on the outcome in patients presenting with an acute coronary syndrome: a cross-sectional study. *BMJ Open* 4: e005077.

Samnegård, A., Silveira, A., Lundman, P., Boquist, S., Odeberg, J., Hulthe, J., McPheat, W., Tornvall, P., Bergstrand, L., Ericsson, C.-G.*, et al.* (2005). Serum matrix metalloproteinase-3 concentration is influenced by MMP-3 -1612 5A/6A promoter genotype and associated with myocardial infarction. *J Intern Med* 258: 411-419.

Schwenk, J.M., Igel, U., Neiman, M., Langen, H., Becker, C., Bjartell, A., Ponten, F., Wiklund, F., Grönberg, H., Nilsson, P.*, et al.* (2010). Toward next generation plasma profiling via heat-induced epitope retrieval and array-based assays. *Mol Cell Proteomics* 9: 2497-2507.

Sjoberg, R., Sundberg, M., Gundberg, A., Sivertsson, A., Schwenk, J.M., Uhlen, M., and Nilsson, P. (2012). Validation of affinity reagents using antigen microarrays. *N Biotechnol* 29: 555-563.

Tsiatsiani, L., and Heck, A.J. (2015). Proteomics beyond trypsin. *FEBS J* 282: 2612-2626.