Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility

Tanja Zeller1, Philipp Wild1, Silke Szymczak2, Maxime Rotival3, Arne Schillert2, Raphaele Castagne3, Seraya Maouche3, Marine Germain3, Karl Lackner4, Heidi Rossmann4, Medea Eleftheriadis1, Christoph R. Sinning1, Renate B. Schnabel1, Edith Lubos1, Detlev Mennerich5, Werner Rust3, Claire Perret3, Carole Proust3, Viviane Nicaud3, Joseph Loscalzo6, Norbert Hübner7, David Tregouet3, Thomas Münnzel1, Andreas Ziegler2, Laurence Tiret3, Stefan Blankenberg1* and François Cambien3*

1 Medizinische Klinik und Poliklinik, Johannes-Gutenberg Universität Mainz, Mainz, Germany, 2 Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätshospital Schleswig-Holstein, Lübeck, Germany, 3 INSERM UMR 937, Pierre et Marie Curie University and Medical School, Paris, France, 4 Institut für Klinische Chemie und Laboratoriumsmedizin, Johannes-Gutenberg Universität Mainz, Mainz, Germany, 5 Boehringer Ingelheim Pharma GmbH and Co. KG, Biberach, Germany, 6 Department of Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, United States of America, 7 Max Delbrück Center for Molecular Medicine, Berlin, Germany

Abstract

**Background:** Variability of gene expression in human may link gene sequence variability and phenotypes; however, non-genetic variations, alone or in combination with genetics, may also influence expression traits and have a critical role in physiological and disease processes.

**Methodology/Principal Findings:** To get better insight into the overall variability of gene expression, we assessed the transcriptome of circulating monocytes, a key cell involved in immunity-related diseases and atherosclerosis, in 1,490 unrelated individuals and investigated its association with >675,000 SNPs and 10 common cardiovascular risk factors. Out of 12,808 expressed genes, 2,745 expression quantitative trait loci were detected ($P<5.78 \times 10^{-15}$), most of them (90%) being cis-modulated. Extensive analyses showed that associations identified by genome-wide association studies of lipids, body mass index or blood pressure were rarely compatible with a mediation by monocyte expression level at the locus. At a study-wide level ($P<3.9 \times 10^{-7}$), 1,662 expression traits (13.0%) were significantly associated with at least one risk factor. Genome-wide interaction analyses suggested that genetic variability and risk factors mostly acted additively on gene expression. Because of the structure of correlation among expression traits, the variability of risk factors could be characterized by a limited set of independent gene expressions which may have biological and clinical relevance. For example expression traits associated with cigarette smoking were more strongly associated with carotid atherosclerosis than smoking itself.

**Conclusions/Significance:** This study demonstrates that the monocyte transcriptome is a potent integrator of genetic and non-genetic influences of relevance for disease pathophysiology and risk assessment.

Citation: Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, et al. (2010) Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility. PLoS ONE 5(5): e10693. doi:10.1371/journal.pone.0010693

Editor: Zoltán Bochdanovits, VU University Medical Center and Center for Neurogenomics and Cognitive Research, VU University, Netherlands

Received March 15, 2010; Accepted April 26, 2010; Published May 18, 2010

Copyright: © 2010 Zeller et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The Gutenberg Heart Study is funded through the government of Rheinland-Pfalz (Stiftung Rheinland Pfalz für Innovation, contract number AZ 961-386261/733), the research programs Wissen schafft Zukunft and Schwerpunkt Vaskuläre Prävention of the Johannes Gutenberg-University of Mainz and its contract with Boehringer Ingelheim and PHILIPS Medical Systems including an unrestricted grant for the Gutenberg Heart Study. Specifically, the research reported in this article was supported by the National Genome Network NGFNplus (contract number project A3 01GS0833) by the Federal Ministry of Education and Research, Germany. For this particular research paper, Boehringer Ingelheim provided payment of two employees for expression microarray analyses and array purchase. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: In this study, Boehringer Ingelheim provided payment for two employees for expression microarray analyses and array purchase and PHILIPS Medical Systems provided instruments for ultrasonic studies. However, this does not alter the authors’ adherence to all the PLoS ONE policies on sharing data and materials, as detailed online in the guide for authors of the Journal.

* E-mail: blankenberg@2-med.klinik.uni-mainz.de (SB); francois.cambien@upmc.fr (FC)

† These authors contributed equally to this work.

Introduction

The transcriptome, i.e. the whole set of RNA transcripts in a cell, is generally conceived as a system whose major function is to pass information encoded in the genome sequence to the realm of phenotypes that underlie physiological and pathological traits. This messenger paradigm justifies the current interest for the genomics of gene expression [1–6] which has been further enhanced by the numerous associations between genetic markers and diseases reported in recent genome-wide association studies (GWAS) and the expected relevance of genome wide expression (GWE) to characterize the biological basis of these associations [7–10]. However the variability of gene expression not only reflects genetic variation but depends on other factors as well such as...
as environmental exposures [11,12], metabolic conditions [13], ageing [14,15] or gender [16–17].

Based on these premises we reasoned that if the state of the transcriptome and its changes are important determinants of cell functions, differences in transcript abundance whatever their origin, genetic or non genetic, may contribute to disease pathogenesis. Moreover, the transcriptome might integrate information from numerous sources and inform on the current pathophysiological state of the organism. To assess these possibilities, a global characterization of the variability of the transcriptome, integrating genetic and non genetic influences was undertaken. The study focused on peripheral blood monocytes because these cells may be isolated from an easily accessible tissue and play a key role in the pathogenesis of immune disorders and atherosclerosis-related diseases [18]. In addition, working on a single cell type reduces the complexity of transcriptome data and may avoid possible biases resulting from the heterogeneous cell-type distribution in different samples as it is the case when using whole blood or leucocytes RNAs.

Results

The genome-wide expression of circulating monocytes

To reduce potential artefacts, fresh samples were collected and processed in a short period of time according to a very strict protocol. Monocytes were obtained from 1,490 unrelated individuals, 730 women and 760 men, aged 35 to 74 years, recruited in the Gutenberg Heart Study (GHS), a community-based project conducted in a single centre in the region of Mainz (Germany) (Table 1). GWE profiles were generated using Illumina Human HT-12 expression BeadChips, and after normalization and filtering out genes undetected in monocytes or non-well characterized (see Materials and Methods), 12,808 expressions traits (averaged over probes) remained for analysis.

Identification of eSNPs and eQTLs

Genotyping was performed using Affymetrix 6.0 arrays. After filtering out SNPs poorly performing or having a minor allele frequency <0.01, 675,350 SNPs were kept for further analyses. All associations between SNPs and expression traits with a P-value $<10^{-3}$ (n=225,000) were stored in the “GHS_Express” database (“GHS_Express” is available online, see Methods S1). At a study-wise threshold of significance correcting for the number of SNPs and expressions ($P<5.78\times10^{-12}$), 37,403 associations, involving 29,912 SNPs and 2,745 expression traits (referred to as eSNPs and eQTLs, respectively), were identified (Table 2). The median number of eSNPs by eQTL was 11 with an interquartile range of 4 to 26. Owing to its large sample size, the study had an 80% power to detect a SNP effect accounting for 4% or more of the variability ($R^2$) of any expression trait. Among the 2,745 significant eQTLs, the $R^2$ observed for the best eSNP varied from 3.1% to >80% with a median of 7.7%. For 290 eQTLs, the $R^2$ was greater than 25%.

Cis versus trans associations

Associations involving SNPs located within 1 Mb of either the 5’ or 3’ end of the associated gene were considered cis (File S1) and other associations were considered trans. In accordance with previous results [2–6], most of the genetic variability affecting the transcriptome was of cis origin. At study-wise significance, the number of cis and trans eQTLs were 2,477 and 349, respectively, yielding a cis/trans ratio of 7.1 (81 eQTLs were both cis- and trans-modulated). At less stringent levels of significance the number of trans associations considerably increased, as expected by chance, whereas the number of cis eQTLs only modestly increased, indicating that the high stringency used for cis eQTLs identification did not result in an important under-estimation of the true number of cis eQTLs (Figure 1).

Comparison with previous GWAS of gene expression

We examined the overlap between the cis eQTLs identified in the present study and those found in three previous association studies in which gene expression was explored in LCLs [1,2] and hepatic cells [3]. For this comparison, a significance threshold of 3.9$x10^{-5}$ (Bonferroni-corrected for 12,808 genes) was used for the analysis of eQTLs in GHS data, corresponding to a single hypothesis tested per gene. Among the cis eQTLs considered significant in each of the studies and involving expression traits detected in GHS, 66.7%, 56.5% and 54.1%, respectively, were significant in our data (Table 3, Files S2–S4). The proportion of cis eQTLs that replicated in GHS increased with the increasing level of significance reported in each study, consistent with the fact that stronger associations are more robust and more likely to be shared by different types of cells. These comparisons revealed a relatively high rate of replication of the previous findings in GHS. However, as a consequence of its greater power, P-values observed in GHS were considerably lower than those previously reported (Figure 2).

We also examined the overlap between cis eQTLs in GHS and cis-heritable eQTLs found by expression profiling of lymphocyte RNA in the San Antonio Family Heart Study (SAFHS) [4]. Among the eQTLs with a cis heritability >0.1 in SAFHS, 62% were significantly cis modulated in GHS, and this proportion reached 89% for heritabilities $\geq0.6$ (Figure 3, File S5).

Trans associations showed much weaker consistency across studies. Among the 50 eQTLs having a trans lod score >4.0 in SAFHS [4] with corresponding expression detected in GHS, only MAPK3P1, was replicated in GHS ($P<10^{-300}$). Replication of the trans associations in studies of similar power as the present one would be of interest.

Identifying eQTLs that may result from the presence of SNPs in probe sequences

For all probes present on the Illumina HT12 array, a systematic search for sequence polymorphisms was undertaken, using the HapMap database as reference (Release 27; Phase II+III, Feb09,
on NCBI B36 assembly and dbSNP b126). Among the 2,477 genes whose expression was associated with \( \text{cis} \) eSNPs, 173 (7\%) were probed by one or several polymorphic sequences (180 probes) (Table S1). For 32 of these probes, the HapMap SNP was present on the Affymetrix array used in this study and for 41 other probes, the HapMap SNP had one or several perfect proxies on the array. For those eQTLs, we cannot exclude the possibility of an artefactual association due to a differential binding of the probe to its target sequence.

**Gene expression, a link between DNA sequence variability and clinical phenotypes?**

A link between genetic variability and clinical phenotypes is supported in human studies by several observations relating variants

### Table 2. Number of gene expression-by-SNP associations at various levels of significance.

| Significance level | Minimum \( R^2 \) | Total number of associations | \( \text{cis/trans} \) ratio for associations | Total number of associated expressions (eQTLs) | \( \text{cis/trans} \) ratio for eQTLs | Total number of associated SNPs (eSNPs) | \( \text{cis/trans} \) ratio for eSNPs |
|-------------------|----------------|-----------------------------|---------------------------------------------|---------------------------------------------|-----------------------------------------|----------------------------------------|-----------------------------------------|
| \(< 10^{-6}\)     | 0.016          | 93491                       | 2.1                                        | 8575                                        | 0.5                                     | 67190                                  | 2.4                                     |
| \(< 10^{-8}\)     | 0.022          | 54749                       | 7.3                                        | 3857                                        | 3.0                                     | 41425                                  | 11.2                                    |
| \(< 10^{-10}\)    | 0.028          | 42421                       | 9.8                                        | 2998                                        | 6.0                                     | 33339                                  | 16.3                                    |
| \(< 5.78 \times 10^{-12}\) | 0.031      | 37403                       | 10.7                                       | 2745                                        | 7.1                                     | 29912                                  | 17.1                                    |
| \(< 10^{-15}\)    | 0.042          | 27330                       | 12.7                                       | 2180                                        | 9.5                                     | 22591                                  | 17.8                                    |
| \(< 10^{-20}\)    | 0.057          | 19655                       | 14.7                                       | 1725                                        | 12.8                                    | 16883                                  | 19.2                                    |
| \(< 10^{-25}\)    | 0.071          | 15015                       | 16.4                                       | 1429                                        | 16.2                                    | 13045                                  | 21.5                                    |
| \(< 10^{-35}\)    | 0.099          | 9673                        | 17.1                                       | 1031                                        | 21.6                                    | 8516                                   | 22.9                                    |
| \(< 10^{-50}\)    | 0.140          | 5873                        | 14.0                                       | 712                                         | 28.8                                    | 5224                                   | 21.7                                    |
| \(< 10^{-100}\)   | 0.263          | 1790                        | 10.5                                       | 290                                         | 28.1                                    | 1598                                   | 11.1                                    |
| \(< 10^{-150}\)   | 0.371          | 922                         | 5.5                                        | 156                                         | 21.4                                    | 772                                    | 5.9                                     |
| \(< 10^{-200}\)   | 0.463          | 635                         | 3.7                                        | 97                                          | 15.3                                    | 504                                    | 3.9                                     |
| \(< 10^{-300}\)   | 0.606          | 321                         | 1.7                                        | 38                                          | 11.7                                    | 213                                    | 1.7                                     |

\(^1\text{Minimum } R^2 \text{ (proportion of gene expression variability explained by a SNP) observed for a given significance level. Numbers corresponding to study-wise significance are shown in bold. For investigating } \text{cis} \text{ associations or performing any other hypothesis-based test, lower levels of significance may be considered.}

doi:10.1371/journal.pone.0010693.t002

---

**Figure 1. Number of eQTLs according to the significance threshold adopted and corresponding \( \text{cis/trans} \) eQTL ratio.** The vertical arrow indicates the study-wise level of significance correcting for the number of hypotheses tested. Some eQTLs being associated with both \( \text{cis} \) and \( \text{trans} \)-acting eSNPs, the sum of \( \text{cis} \) and \( \text{trans} \) eQTLs is greater than the total number of eQTLs.

doi:10.1371/journal.pone.0010693.g001
Figure 2. Comparison of the distributions of \( \mathcal{P} \)-values of \( \text{cis} \) eQTLs reported as significant in three previous association studies with \( \mathcal{P} \)-values observed in GHS for the same eQTLs. For each of the 3 comparisons, we selected in GHS the subset of gene expressions claimed as significant in the study of comparison. Only autosomal genes were considered in these comparisons. The data used to generate this figure are provided in Files S2–S4. See also footnote of Table S3 for details.

doi:10.1371/journal.pone.0010693.g002

Table 3. Number of \( \text{cis} \) eQTLs identified in previous studies and replicated in GHS.

| Level of significance | Stranger et al. | Dixon et al. | Schadt et al. |
|-----------------------|-----------------|--------------|--------------|
| \( >10^{-8} \)        | 86              | 110          | 928          |
| \( 10^{-9} - 10^{-10} \) | 63              | 162          | 168          |
| \( 10^{-11} - 10^{-12} \) | 144            | 237          | 211          |
| \( 10^{-12} - 10^{-13} \) | 60              | 102          | 120          |
| \( 10^{-13} - 10^{-14} \) | 15              | 144          | 237          |
| \( 10^{-14} - 10^{-15} \) | 20              | 60           | 102          |
| \( 10^{-15} - 10^{-16} \) | 38              | 89.5         | 73           |
| \( 10^{-16} - 10^{-17} \) | 48              | 89           | 89           |
|\( \leq 10^{-25} \) | 439             | 773          | 1603         |

* Comparisons were based on sets of gene expressions overlapping between each study and GHS and were restricted to autosomal \( \text{cis} \) eQTLs. All \( \text{cis} \) eQTLs considered significant in each study were retrieved and replication was assessed in GHS (\( \mathcal{P}<3.9\times10^{-6} \) correcting for 12,808 gene expressions).

For Stranger et al [1], data were extracted from Table S2. We considered as significant the associations found in at least 3 HAPMAP populations. For Dixon et al [2], data were extracted from Table S1 and \( \text{trans} \) eQTLs were excluded. Matching of probes was done using a table provided by the authors on their web site. For Schadt et al [3], \( \text{cis} \) eQTLs considered significant (First.Pass.Indicator set to 1) were extracted from Table S3. For each eQTL, we selected in GHS the \( \mathcal{P} \)-value of the best \( \text{cis} \) eSNP. The full data used to generate this table are provided in Files S2–S4.

doi:10.1371/journal.pone.0010693.t003
in regulatory gene regions to protein phenotypes or diseases [19,20]. However, this message-passing paradigm has never been evaluated on a genome-wide basis. We therefore tested whether monocytes gene expression might mediate the effects of loci recently identified by GWAS of cardiovascular risk factors. For each locus identified in GWAS of lipid variables [21], blood pressure (BP) [22] and body mass index (BMI) [23], we selected the lead SNP or a tag SNP having an $r^2 \geq 0.8$ with the lead SNP in GHS data. Associations between lead/tag SNPs and corresponding risk factors were checked in all GHS subjects for whom genome-wide data was available ($n = 3,175$).

Most previous GWAS loci for circulating lipids were replicated in our data (Table 4) but only few of the findings of GWAS of BMI and BP were replicated (Table S2). This low replication is probably due to a lack of power, as the maximum $R^2$ observed in the GWAS of BP [22] was 0.09% and the power of GHS to replicate such an association was only 38%.

For each GWAS locus, we examined whether the lead/tag SNP correlated with any expression trait in GHS data and when a significant association was found, we checked whether the expression trait was significantly associated with the risk factor under consideration. This analysis revealed that very few GWAS results were compatible with an effect mediated by gene expression at the locus (Table 4 and Table S2). There were, however, two exceptions: the first one concerned the $LPL$ locus, where the minor allele of rs17489282 was associated with higher HDL-cholesterol ($P = 5.91 \times 10^{-25}$) and $LPL$ expression ($P = 2.18 \times 10^{-29}$), while HDL-cholesterol and $LPL$ expression were positively correlated ($P = 6 \times 10^{-24}$), consistent with an effect mediated by LPL; the second one concerned the association between the 1p13.3 locus and LDL-cholesterol. This locus encompasses three potential candidate genes, $CELSR2$, $PSRC1$ and $SORT1$, and it has been suggested that $CELSR2$ or $SORT1$ could be responsible for the reported associations of this locus with LDL [3,24,25]. In our data, the minor allele of rs629301 (a perfect tag of the lead SNP identified by GWAS), was associated with lower LDL-cholesterol ($P = 2.6 \times 10^{-24}$) and higher $PRSC1$ expression ($P = 2.3 \times 10^{-36}$).
while PRSCI expression and LDL-cholesterol were negatively correlated (P = 0.019). Results for CELSR2 were much less consistent and SORT1, the third gene at the locus, was not cis-modulated in monocytes.

Several loci associated with coronary artery disease (CAD) have been identified by GWAS [26–29]. The strongest association involves SNPs in the 9p21 region. Recently it was reported that deletion in mice of the region orthologous to the 9p21 CAD interval in humans affects the expression of the nearby CDKN2a and cdkn2b genes as well as the properties of proliferation of vascular cells [30]. The Cyclin-dependent kinase inhibitor coding genes, CDKN2a and CDKN2b, are also located close to the CAD locus in humans. CDKN2a expression in monocytes was not detected in our study, we therefore focused our analysis on CDKN2b. All SNPs available in GHS in the region encompassing the CAD locus were tested for association with expression of CDKN2b. Figure 4 shows that CDKN2b expression was strongly associated with several SNPs located in a region upstream of the gene sequence (P<10^-46). However, these SNPs were not associated with CAD (this result was obtained in a yet unpublished GWAS comparing GHS individuals to a cohort of CAD patients), whereas proxies of the CAD-associated SNPs were unrelated with CDKN2b expression (see legend of Figure 4 for more details). The SNPs associated with CDKN2b expression are located within the sequence of the non-coding alternatively spliced gene ANRIL (also named CDKN2BAS) whose implication in the association with CAD has been hypothesized [31]. Although our results are limited by the fact that neither CDKN2b nor ANRIL expressions could be

**Table 4. Loci identified in GWAS of circulating lipids – associations of lead/tag SNPs with phenotypes and expression, and of expression with phenotype in GHS.**

| Lead SNP in GWAS | Phenotype | Chr | Position (Mb) | Genes in region | Tag SNP in Affy 6.0 with r^2>0.8 | r^2 between lead SNP and tag SNP | Association between tag SNP and phenotype (P-value) | eQTL associated with tag SNP | Association between tag SNP and eQTL (P-value) | Association between eQTL and phenotype (P-value) |
|------------------|-----------|-----|--------------|----------------|-----------------|----------------------|-----------------------------|-----------------|-------------------------|-----------------------------|
| rs10889353       | TG        | 1   | 62.83        | DOCK7          | rs10889353      | 1.00                 | 1.86E-04                   | DOCK7           | 7.09E-52                | 0.3970                      |
| rs646776*        | LDL       | 1   | 109.53       | CELSR2/PSRC1/  SORT1 | rs629301      | 1.00                 | 2.62E-04                   | PSRC1           | 2.34E-56                | 0.0190                      |
| rs693            | LDL       | 2   | 21.14        | APO8           | rs693           | 1.00                 | 1.61E-04                   | none            | none                    | none                        |
| rs6754295        | HDL       | 2   | 21.12        | APO8           | rs673548        | 0.86                 | 0.0435                     | none            | none                    | none                        |
| rs673548         | TG        | 2   | 21.15        | APO8           | rs673548        | 1.00                 | 4.10E-05                   | none            | none                    | none                        |
| rs780094         | TG        | 2   | 27.65        | GCKR           | rs780094        | 1.00                 | 3.15E-08                   | none            | none                    | none                        |
| rs6756629        | LDL       | 2   | 43.98        | ABCG5          | rs4953023       | 1.00                 | 7.87E-05                   | none            | none                    | none                        |
| rs3846662        | LDL       | 5   | 74.69        | HMGCR          | rs12654264      | 0.84                 | 9.26E-06                   | none            | none                    | none                        |
| rs12670798       | LDL       | 7   | 21.38        | DNAH11         | none            |                      | none                        | none            | none                    | none                        |
| rs2240466        | TG        | 7   | 72.39        | MLXPL          | rs2074755       | 1.00                 | 0.0015                     | none            | none                    | none                        |
| rs2083637*       | HDL       | 8   | 19.91        | LPL            | rs17489282      | 1.00                 | 5.91E-05                   | LPL             | 2.18E-06                | 0.0006                      |
| rs2083637        | TG        | 8   | 19.91        | LPL            | rs17489282      | 1.00                 | 3.31E-07                   | LPL             | 2.18E-06                | 0.3520                      |
| rs3905000        | HDL       | 9   | 104.74       | ABCA1          | rs3890182       | 1.00                 | 0.33                      | none            | none                    | none                        |
| rs7395662        | HDL       | 11  | 48.48        | MADD-FOLH1     | rs7395662       | 1.00                 | 0.17                      | MYBP3           | 1.17E-09                | 0.1435                      |
| rs174570         | LDL       | 11  | 61.35        | FADS2/3        | rs174570        | 1.00                 | 0.0386                     | none            | none                    | none                        |
| rs12272004       | TG        | 11  | 116.11       | APOA1(A4/A5/C3)| rs10488699      | 1.00                 | 1.63E-06                   | none            | none                    | none                        |
| rs1532085        | HDL       | 15  | 56.47        | LPC            | none            |                      | none                        | none            | none                    | none                        |
| rs1532624        | HDL       | 16  | 55.56        | CETF           | none            |                      | none                        | none            | none                    | none                        |
| rs2271293        | HDL       | 16  | 66.46        | CTCF-PRMT8     | rs2271293       | 1.00                 | 0.0145                     | DPEP3           | 5.09E-17                | 0.8275                      |
| rs4939883        | HDL       | 18  | 45.42        | LIPC           | rs7240405       | 1.00                 | 0.0233                     | none            | none                    | none                        |
| rs2228671        | LDL       | 19  | 11.07        | LDLR           | none            |                      | none                        | none            | none                    | none                        |
| rs157580         | LDL       | 19  | 50.09        | TOMM40-APOE    | none            |                      | none                        | none            | none                    | none                        |
evaluated, they reveal that in humans, SNPs that affect CDKN2B expression are different from those that are known to affect CAD risk (Figure 4).

Expression traits associated with risk factors

To investigate gene expression in relation to risk factors (age, gender, BMI, HDL and LDL cholesterol, triglycerides, Systolic and Diastolic Blood pressure, smoking and plasma CRP), the study-wise significance threshold was set at $3.9 \times 10^{-7}$ to correct for the number of risk factors ($n = 10$) and expressions ($n = 12,808$) tested. Overall, 1,662 expression traits (13.0%) were associated with at least one risk factor (Table 5 and File S6). Gender and age were the two major factors influencing expression levels (807 gene expressions were affected by gender and 396 by age). BMI, smoking and C-reactive protein (CRP) levels were also correlated with numerous expression traits (230, 294 and 328, respectively). Conversely, few associations with BP and lipids were observed (Table 5).

Genetic and non-genetic factors act additively on gene expression

Cis-eQTLs were over-represented among expression traits that were also affected by gender, age, BMI, smoking and CRP, with odds ratios as high as 3.24 for cigarette smoking (Table 5). This suggests that some genes are more responsive than others to the influence of multiple factors. For expression traits that were simultaneously associated with cis-eSNPs and risk factors ($n = 465$), we determined the joint effects of the two sources of variability on expression level. For this purpose, each eQTL was modelled as a function of the best cis-acting eSNP, the associated risk factor and
Table 5. Number of expression traits associated with risk factors and cis eSNPs.

| Risk factor | Number of expression traits associated with the specified risk factor | Number (%) of expression traits also associated with cis eSNPs | Odds ratio (95% CI)* |
|-------------|---------------------------------------------------------------------|----------------------------------------------------------|---------------------|
| Gender      | 807                                                                 | 230 (28.3%)                                              | 1.73 (1.47–2.03)    |
| Age         | 396                                                                 | 94 (23.7%)                                               | 1.31 (1.03–1.66)    |
| BMI         | 230                                                                 | 72 (31.3%)                                               | 1.92 (1.45–2.56)    |
| HDL         | 9                                                                    | 1 (11.1%)                                                | ND                  |
| LDL         | 1                                                                    | 0                                                       | ND                  |
| Triglycerides | 9                                                                 | 3 (33.3%)                                                | ND                  |
| SBP         | 48                                                                   | 6 (12.5%)                                                | 0.59 (0.25–1.40)    |
| DBP         | 18                                                                   | 2 (11.1%)                                                | ND                  |
| Smoking     | 294                                                                  | 126 (42.9%)                                              | 3.24 (2.56–4.10)    |
| CRP         | 328                                                                  | 116 (35.4%)                                              | 2.34 (1.86–2.95)    |
| All (irrespective of any association with risk factor) | 12,808 | 2,477 (19.3%) |

Study-wise levels of significance were considered for associations of expression traits with risk factors and SNPs (3.9 × 10⁻⁷ and 5.78 × 10⁻¹², respectively). Associations of expression traits with BMI, CRP and smoking were adjusted for age and sex, and association with HDL, LDL, triglycerides, SBP and DBP were additionally adjusted for BMI.

*Odds ratio (OR) of being influenced by a cis eSNP for an expression trait associated with a given risk factor. For example, gender-related expression traits have an OR of 1.73 of being influenced by cis eSNPs by comparison to expression traits unrelated to gender. ND: not determined because of small numbers.

doi:10.1371/journal.pone.0010693.t005

The Monocytes Transcriptome

the interaction between the two. When several risk factors were associated with the same eQTLs, each of them was tested separately for interaction with the corresponding SNP. The best FDR-corrected P-values for interaction (File S7) were 0.014 (JSCU expression, gender and rs4830487) and the second one was 0.042 (HIST1H2AE expression, smoking and rs16891378). This first genome-wide exploration of interaction between cis eSNPs and risk factors on gene expression therefore suggests that the two sources of variability mostly act additively on expression. This is illustrated for eQTLs affected by smoking in Figure 5. It must be noted however that despite the large size of this study, its power may nevertheless be insufficient to assess weak interaction.

Expressions influenced by genetic and risk factors are enriched in immunity and defense genes

An ontology analysis using the Panther system demonstrated that, by reference to the list of 12,808 genes expressed in monocytes, the set of 465 expression traits affected by multiple sources of variability was enriched in “Immunity and defense” genes (69 observed/35.8 expected, P = 4.5 × 10⁻⁶), especially in the sub-categories of “Macrophage-mediated immunity” (18/3.6, P = 7.5 × 10⁻⁶).

The variability of each risk factor can be characterized by a limited set of independent gene expressions

The preceding analyses revealed that each risk factor was associated with a large number of expression traits, thus emphasizing the multiple inter-relations existing between the transcriptomic and risk factor profiles of an individual. While it is important from a biological and mechanistic perspective to characterize at best all the genes that are influenced by a given condition, from a clinical perspective, it might be more relevant to identify a limited set of gene expressions that could efficiently discriminate individuals with different risk profiles. Indeed, because of the tight co-regulation of genes within biological systems, numerous gene expressions are inter-correlated and consequently, their association with risk factors are not independent. To account for this inter-dependency, we conducted a multivariate analysis to identify expression traits that were independently associated with each risk factor.

To obtain reliable results, we randomly divided the study population into two sub-samples of equal size which were used for screening and validation purposes respectively (see Materials and Methods). In these analyses, each risk factor was considered separately whereas all expression traits were envisaged jointly for their potential association with the risk factor, considered here as the dependent variable. The screening/validation procedure was repeated 250 times and for each risk factor, we report expression traits associated (P < 0.01) with the risk factor in more than 25% of the replicates. This stringent approach led to the identification of 106 independent expression correlates for the ten risk factors (Table 6), a much reduced number compared to the 1,662 expression traits previously identified by the one-to-one association analysis presented in Table 5.

Gender and age. Even after exclusion of sex-linked genes, gender was independently associated with the largest number of expression traits (n = 31) which, considered all together, contributed to a highly significant discrimination between males and females (P < 10⁻¹⁰). By contrast the number of expression traits independently associated with age was more limited (n = 12).

BMI and CRP. Both factors were independently associated with 12 and 14 expression traits respectively. Inspection of the genes listed in Table 6 shows that several of them, including CX3CR1, CD209, CLEC10A, FCER1A, FCGBP, C1RL, C1QB, CD36, ADM and VSIG4, encode proteins involved in the differentiation or maturation of immunity-related cells and in host defence [32–38]. We may speculate that the variability of expression of these genes is the consequence of an already present heterogeneity of monocytes [39,40] or reflects a particular transcription pattern that prefigures future functional changes. The example of CX3CR1 which was positively associated with both BMI and CRP is particularly interesting as this gene encodes the fractalkine receptor whose role is essential in the migration of
monocytes to sites of inflammation and injury, especially in atherosclerotic lesions [41].

**Lipids.** ABCA1 and ABCG1 gene expressions were both associated with circulating lipids. The proteins encoded by these genes are key players in reverse cholesterol transport and the regulation of lipid-trafficking mechanisms in macrophages respectively [42,43]. MYLIP (Idol) is a ligase involved in the ubiquitination and degradation of LDL receptors [44,45] and SCD is a stearoyl-CoA desaturase involved in the conversion of saturated into monounsaturated fatty acids that regulates lipid metabolism and may be modulated by dietary intake [46].

**Blood pressure.** One of the main independent correlates of SBP was ARID5B (MRF2), whose relevance in the physiology of BP regulation is supported by its role as a regulator of smooth muscle differentiation and proliferation [47]. GFODI1, another expression correlate of SBP and DBP is a gene of unknown function which has been associated with attention deficit hyperactivity disorder [48].

Cigarette smoking has a major impact on gene expression and atherosclerosis

Smoking was independently associated with 18 expression traits (Table 6) which, considered all together, contributed to a highly significant discrimination between smokers and non-smokers (P < 10^{-10}, R^2 > 50%). Nine of these genes were modulated by cis eSNPs (Table 7) and as already mentioned above, the genetic and smoking effects on these gene expressions were additive (Figure 5). Among the 18 expression traits associated with smoking, SASH1, P2RY6 and PTGDS were systematically retrieved in all
screening/validation replicates (Table S3). *SASH1* is a tumor suppressor gene [49], *P2RY6* encodes a G-protein-coupled receptor involved in the proinflammatory response to UDP in monocytes [50] and *PTGDS* encodes a prostaglandin D synthase involved in smooth muscle contraction/relaxation and inhibition of platelet aggregation, two functions known to be modified by tobacco consumption. Recently, in a GWE study of leukocytes from atherosclerotic plaques after adjustment for smoking status may indicate a broader association was detected (Table 7).

**Discussion**

This large-scale investigation of the transcriptome of monocytes in healthy individuals provides new biological insights into the mechanisms by which gene expression might contribute to disease pathogenesis. In the line of previous studies [1–5], we could build a detailed map of cis-regulated eQTLs in monocytes. Even if cell-specific eQTLs exist [53], a large fraction of them are likely to be common to other cell types, and the eQTL map provided here constitutes the most extensive one so far.

Despite the large number of eQTLs identified, the transcriptome of circulating monocytes, contrary to initial expectations [7–10], appeared of modest help to dissect the relationship between genome variability and complex human traits such as cardiovascular risk factors. One explanation for this finding might be that monocytes are not the most relevant cells for unravelling links between genome variation and the risk factors investigated. With regard to circulating lipids for example, only 28 of the 45 genes located in regions harbouring SNPs associated with circulating lipids in GWAS [24] were expressed in monocytes. The difficulty to corroborate the messenger paradigm in human clinical studies may also relate to the fact that the linear model of the dichotomous variable used to define smoking. This illustrates the dual aspect of the transcriptome which may be viewed either as an element in a causal chain or as reflecting ongoing processes with no implied causation. Because genetics may help to dissect causal pathways, we examined whether the best cis-eSNPs associated with expression of the smoking-related genes were also associated with carotid atherosclerosis, but no such association was detected (Table 7).

**Table 6.** Subsets of expression traits showing robust independent association with the different risk factors in the validation sample.

| Risk factor | Median (range) of the global P-values across replicates* | List of gene expressions associated with risk factor after adjustment on covariates† |
|-------------|--------------------------------------------------------|-----------------------------------------------------------------------------------|
| Gender      | All, P<10^{-10}                                        | CCDC106, MMEL1, ANKRDS7, CXCR7, FCGR2B, SOX15, FCGBP, LPXN, CD24, HOXA9, PTK6, DDX43, RAB11F1P1, PK2, CLEC4G, ADAR1B, PKR2, MYBPH1, PER5, TPRP3, MPO, FAM24B, EMR3, ENOSF1, TPM2, PTGDS, COL5R3, CD1A, FOLR2, BOLAS, OPLAH |
| Age         | 1.1x10^{-50} (2.9x10^{-29}–3.6x10^{-37})             | PARP3, PDGFRB, NEFH, P2RY2, SPINK2, GPER, NFKB1G, ZSCAN18, IGLL1, BLK, ITM2C, C1RL |
| BMI         | 1.5x10^{-57} (2.6x10^{-35}–2.7x10^{-26})             | CX3CR1, MAP3K6, FCGBP, CD209, LYPD2, VSG4A, RPRG1P1, PACAP, LGALS3BP, ELA2, CD36, ABCA1 |
| HDL-chol    | 5.5x10^{-40} (2.2x10^{-35}–2.7x10^{-3})              | PRDM1, SCD, DPEP2, TMEM43 |
| LDL-chol    | 2.3x10^{-39} (3.8x10^{-37}–0.5)                       | BCSL, ABCA1 |
| Triglycerides | 4.2x10^{-6} (2.1x10^{-11}–0.08)                        | MYLIP, PHGDH, ABCA1, ELA2, ABGC1, SASH1 |
| SBP         | 5.0x10^{-20} (4.6x10^{-27}–2.6x10^{-12})             | CRIP1, GOFD1, DHR59, NR4A2, TSC22D3, ARID5B, PAPPS2, HVNC1 |
| DBP         | 1.8x10^{-10} (3.6x10^{-12}–1.2x10^{-4})              | GOFD1, CRIP1, TPRP3, NR4A2,EMPL1 |
| CRP         | 6.1x10^{-41} (7.3x10^{-47}–2.0x10^{-10})             | FAM20A, CEP5, FCGBP, COL9A2, C1RL, ADM, CREBS, APBB1P1, CX3CR1, C1Q8, M544A4A, FCER1A, ALDH1A1, FLVCR2 |
| Smoking     | 9.7x10^{-108} (1.9x10^{-128}–3.7x10^{-94})           | SASH1, P2RY6, PTGDS, PID1, CYP4F22, MMP25, WWC3, FUC1A, PDE4B, STAB1, GFRA2, CLEC10A, CAMKTD2, DHR59, CNTPAP2, IQCK, ITG6B, SMAD6 |

*The global P-value is the P-value obtained by comparing the model with all significant expression traits and covariates to the model with covariates only. Expressions that are underlined are associated negatively to the risk factor (or to male gender), others are associated positively (or to female gender). †Covariates: age and gender for BMI, CRP and smoking; age, gender and BMI for lipids and BP. ‡Gender-associated traits were selected from autosomal genes only.

doi:10.1371/journal.pone.0010693.t006
causality generally assumed to reflect the relation between genome variability, expression and phenotype may be too simplistic to account for a much more complex biological reality. The effects of genetic variants may be too weak to allow detection even in a study of this size. It is also important to keep in mind that most reported eSNPs are acting in cis, whereas trans eSNPs may actually be those that mainly drive the changes in gene expression that affect disease risk.

Most importantly, the present study highlighted for the first time the strong link existing between the transcriptome of an individual and his/her clinical and epidemiological profile. The fact that the transcriptome tightly mirrors the variability of risk factors at a cellular level may have profound implications from a biological and clinical perspective. Until now, the traditional way of viewing the role of genes in the susceptibility to human diseases was through the effect of their variability of sequence. The present findings suggest that another important, if not greater, impact of genes on human phenotypes relates to the variability of their expression, whatever the origin of this variability. The global association observed between most cardiovascular risk factors and the transcriptome and the fact that each risk factor could be characterized by a limited and specific set of independent gene expressions further suggests that this relationship might be clinically relevant. This was particularly well illustrated by the response of the transcriptome to cigarette smoking. We showed that less than 20 genes among the 12,000 expressed in monocytes could highly discriminate smokers and non-smokers, and among them, four genes were sufficient to account for the strong association existing between smoking and atherosclerosis. Whether these genes are causally involved in the mechanisms linking smoking to the development of atherosclerotic plaques or whether they are only markers of ongoing pathological processes remains to be elucidated.

In conclusion, the variability of the transcriptome of monocytes can be viewed from two perspectives. On one hand it reflects the accumulation of effects originating from the genome and the environment and may inform on a number of ongoing processes relevant to disease. On the other hand, it may reflect or anticipate differences in monocytes biology that could have pathophysiological implications. This dual perspective suggests that a better understanding of the sources of variability of the transcriptome of monocytes and other easily accessible cells, will contribute in an important way to our understanding of complex diseases.

### Materials and Methods

**Ethic statement**

The study protocol and drawing of the blood sample have been approved by the local ethics committee and by the local and federal data safety commissioners (Ethik-Kommission der Landesarztekammer Rheinland-Pfalz). All subjects included signed an informed consent.

**Study population**

The Gutenberg Heart Study (GHS) is designed as a community-based, prospective, observational single-center cohort study in the Rhein-Main region in western mid-Germany. The primary aim of the study is to improve the individual cardiovascular risk

### Table 7. Smoking-related expression traits: association of expression with cis-acting eSNP, smoking and extent of atherosclerosis.

| Smoking-related gene expressions | Association of the best cis eSNP with expression | Association of the best cis eSNP with the extent of atherosclerosis | Association of expression with smoking | Association of expression with the extent of atherosclerosis |
|--------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| SNP number | P-value  | t-value | P-value | t-value | P-value | t-value | P-value |
| CAMK1D | rs4478891 | 1.1×10⁻⁶ | 0.7 | 0.51 | −7.4 | 2.3×10⁻⁵ | −3.1 | 0.002 |
| CLEC10A | rs367953 | 3.6×10⁻¹² | 0.5 | 0.65 | 13.0 | 5.4×10⁻⁷ | −0.4 | 0.68 |
| CNTNAP2 | rs1110144 | 8.6×10⁻¹² | −1.1 | 0.28 | 6.5 | 7.3×10⁻¹¹ | 1.7 | 0.082 |
| CYP4F22 | rs11253478 | 1.5×10⁻⁶ | 0.0 | 1.00 | −17.5 | 1.8×10⁻¹² | −2.2 | 0.027 |
| DHR59 | rs1386426 | 4.3×10⁻⁷ | 2.5 | 0.012 | −6.3 | 3.5×10⁻¹⁰ | 0.2 | 0.82 |
| FUC1A | rs9424398 | 6.8×10⁻⁶ | 1.0 | 0.34 | 13.7 | 1.4×10⁻⁹ | 2.1 | 0.035 |
| GFRα2 | rs1479056 | 1.6×10⁻¹⁰⁸ | 0.5 | 0.62 | 13.3 | 3.1×10⁻⁸ | 0.8 | 0.44 |
| IQK | rs1879894 | 1.1×10⁻⁷ | 0.6 | 0.52 | −7.6 | 2.7×10⁻¹⁴ | −1.7 | 0.089 |
| ITGB7 | rs17080239 | 1.0×10⁻⁶ | 0.4 | 0.68 | 6.3 | 3.9×10⁻¹⁰ | 0.3 | 0.73 |
| MMP25 | rs7188573 | 4.7×10⁻¹⁸ | 0.0 | 1.00 | 15.6 | 5.4×10⁻⁵ | 3.6 | 3.51×10⁻⁴ |
| P2RY6 | rs781305 | 9.0×10⁻⁷ | −1.0 | 0.34 | 16.0 | 1.5×10⁻⁵ | 0.5 | 0.58 |
| PDE4B | rs4352802 | 4.5×10⁻⁷ | −1.2 | 0.24 | −7.7 | 1.2×10⁻¹⁴ | −1.8 | 0.077 |
| PIK | rs4972894 | 1.1×10⁻²⁵ | −0.1 | 0.95 | 10.5 | 6.2×10⁻²⁵ | 1.4 | 0.16 |
| PTGDS | rs10870158 | 6.0×10⁻⁹ | 0.7 | 0.49 | −14.0 | 3.5×10⁻⁴ | −5.2 | 1.77×10⁻⁷ |
| SASH1 | rs17712470 | 5.6×10⁻¹⁴ | 0.5 | 0.63 | 20.5 | 5.1×10⁻⁹ | 4.4 | 1.38×10⁻⁵ |
| SMAD6 | rs17264185 | 2.7×10⁻¹⁵ | −1.5 | 0.13 | 6.8 | 9.6×10⁻¹² | 1.6 | 0.11 |
| STAB1 | rs9867823 | 2.8×10⁻²⁸ | 0.4 | 0.68 | 12.3 | 2.9×10⁻³ | 0.7 | 0.50 |
| WWC3 | rs1013478 | 8.0×10⁻⁶ | −1.1 | 0.26 | 10.6 | 1.8×10⁻²⁵ | 3.4 | 6.32×10⁻⁴ |

Atherosclerosis was assessed by the number of carotid plaques. Associations of expressions with smoking and number of carotid plaques were adjusted on gender and age. Cis eSNPs significant at a study-wise level (P<5.78×10⁻¹²) and associations with carotid plaques significant after correction for multiple testing (n = 18 tests) are shown in bold characters.

doi:10.1371/journal.pone.0010693.t007
prediction by identifying genetic and non genetic risk factors contributing to cardiovascular diseases, with a strong emphasis on atherosclerosis.

A sample of eligible participants was randomly drawn from the registers of the local registry offices in the city of Mainz and the district of Mainz-Bingen. This sample was stratified in a ratio of 1:1 for gender and residence, and in equal numbers for decades of age. Inclusion criteria were an age between 35 and 74 years and a written consent; exclusion criteria were insufficient knowledge of the German language to understand explanations and instructions, and physical or psychic inability to participate in the examinations in the study center. Individuals were invited for a 3-hour baseline-examination to the study center where clinical examinations and collection of blood samples were performed. The present analysis was based on an initial sample of 3,336 subjects successively enrolled into the GHS from April 2007 to April 2008. Genomic DNA was isolated from all participants. Monocyte RNA was isolated from half of the participants recruited each day to ensure rapid sample processing and isolation of total RNA. For approximately 1,500 study participants, both DNA and RNA were available.

Measurement and definition of cardiovascular risk factors

Blood pressure measurements were performed by an automated sphygmomanometer blood pressure meter (Omron 705CP-II, OMRON Medizintechnik Handegesellschaft GmbH, Germany) after 5, 8 and 11 minutes of rest. The mean from the 2nd and 3rd standardized measurement was calculated for the systolic and diastolic blood pressure. For the anthropometric measurements, calibrated, digital scales (Seca 220, Seca Germany), a measuring stick (Seca 220, Seca, Germany) and a waist measuring tape were used. The blood sampling was carried out under fasting conditions. HDL-cholesterol, LDL-cholesterol, triglycerides and C-reactive protein (CRP) measurements were performed on an Architect c8000 by commercially available tests (CRP, Ultra C-reactive protein (CRP) measurements were performed on an Architect c8000 by commercially available tests (CRP, Ultra HDL, Direct LDL and Triglycerides) from Abbott (www. abbottdiagnostics.de). All tests were measured under standardized conditions. The blood sampling was carried out under fasting conditions. HDL-cholesterol, LDL-cholesterol, triglycerides and C-reactive protein (CRP) measurements were performed on an Architect c8000 by commercially available tests (CRP, Ultra HDL, Direct LDL and Triglycerides) from Abbott (www. abbottdiagnostics.de). All tests were measured under standardized conditions.

Ultrasound of the Carotid Arteries and evaluation of the number of atherosclerotic plaques

IMT was assessed with an iE33 ultrasound system (Philips, NL) using an 11 to 3 MHz linear array transducer. Experienced technologists blinded to participants’ clinical data made all ultrasound measurements. The IMT was visualized bilaterally at the far wall of the CCA. In brief, a cursor representing the region of interest (10 mm) was positioned 1 cm in front of the beginning before the carotid bulb. Evaluation was performed using an automatic computerized system (Philips, NL, Qlab software) and triggering was performed according to the Q wave of the ECG to enable measurement in complete relaxation of the ventricle. IMT was recorded 1 cm before the carotid bulb in a part without plaque on the left and right side. As mean IMT, the CCA was reported with the sum of IMT of the left and right side and afterwards divided by two. Plaques were defined as thickening of the IMT of at least 1.5 mm and presence was checked in all measured arteries. The number of plaques from both sides was recorded and subjects being classified as plaque positive when at least one plaque was measured on either side or plaque negative, when no plaque was recorded.

Genotyping

For each participant genomic DNA was extracted from buffy-coats prepared from EDTA blood samples (9 mL) using the method of Miller [34]. Genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (http://www. affymetrix.com), as described by the Affymetrix user manual. Genotypes were called using the Affymetrix Birdseed-V2 calling algorithm and quality control was performed using GenABEL (http://mga.bionet.nsc.ru/nlru/GenABEL/). Individuals with a call rate below 97% or a too high autosomal heterozygosity (False Discovery Rate <1%) were excluded. After applying standard quality criteria (minor allele frequency >1%, genotype call rate >98% and P-value of deviation from Hardy-Weinberg equilibrium >10-4), 675,350 out of 900,392 SNPs remained for analysis.

Separation of monocytes

Separation of monocytes was conducted within 60 min after blood collection and RNA was extracted the same day. Total RNA was isolated from monocytes using Trizol extraction and purification by silica-based columns. To separate monocytes, 8 mL blood was collected using the Vacutainer CPT Cell Preparation Tube System (BD, Heidelberg, Germany) and 400 μL, Rosette Sep Monocyte Enrichment Cocktail (StemCell Technologies, Vancouver, Canada) was added immediately after blood collection. Monocytes, not labeled by antibodies, are collected as a highly enriched fraction at the interface between plasma and the density medium in the tube. After separation, cells were washed twice in ice cold PBS buffer containing 2 mM EDTA. Success of monocyte separation was controlled using an ADVIA 2120 Analyser (Siemens Healthcare Diagnostics, Eschborn, Germany) for part of the samples.

Preparation of RNA

After separation, cells were resuspended in 1.5 mL Trizol Reagent (Invitrogen, Karlsruhe, Germany) immediately and frozen at −20°C until isolation of RNA at the same day (maximal storage time 5 h). After thawing, samples were transferred into Phase Lock Gel Tubes (Eppendorf, Hamburg, Germany), 200 mL chloroform was added and phases were separated by centrifugation at 4600 rpm for 15 min. Purification of total monocytes RNA was performed using the RNaseasy Mini kit (Qiagen, Hilden, Germany) according to the manufacturers’ Animal Cell Spin and RNA Cleanup protocols including an additional DNase digestion step. Total RNA was eluted in 20 μL RNase-free water. Yield of RNA was checked spectrophotometrically by NanoDrop N-1000 measuring the OD260 as well as the ratio OD260 and OD280. The integrity of the total RNA was assessed through analysis on an Agilent Bioanalyzer 2100 (Agilent Technologies, Boeblingen, Germany).

Genome-Wide Expression analysis

GWE analysis was performed on monocytes RNA samples using the Illumina HT-12 v3 BeadChip (http://www.Illumina. com). RNA samples were processed in batches of 96 samples. Two hundred ng of total RNA was reverse transcribed, amplified and biotinylated using the Illumina TotalPrep-96 RNA Amplification Kit (Ambion/Applied Biosystems, Darmstadt, Germany). 700 ng of each biotinylated cRNA was hybridized to a single BeadChip at
58°C for 16–18 hours. BeadChips were scanned using the Illumina Bead Array Reader.

**Pre-processing of expression data**

The summary probe-level data delivered by the Illumina scanner (mean and SD computed over all beads for a particular probe) was loaded in Beadstudio. The pre-processing done by the Illumina software, at the level of the scanner and by Beadstudio included: correction for local background effects, removal of outlier beads, computation of average bead signal and SD for each probe and gene, calculation of detection P-values using negative controls present on the array, quantile normalization across arrays, check of outlier samples using a clustering algorithm, check of positive controls. Analyses were carried out on the mean level for all probes in each gene. To stabilise variance across expression levels, we applied an arcsinh transformation to the expression data [53]. Compared to a log transformation, this transformation has the advantage not to discard negative expression values which can occur in Illumina data.

The Illumina HT-12 BeadChip included 37,804 genes (some probes being not assigned to RefSeq genes). A gene was declared significantly expressed in the dataset, i.e. expressed above background (as measured by the negative controls present on each array), when the detection P-value calculated by Beadstudio was ≤0.05 in more than 5% of the samples. This resulted in 22,305 genes considered as being significantly expressed in our dataset. After removing 8,058 putative and/or non well characterized genes, i.e. gene names starting by KIAA (n = 165), FLJ (n = 214), HS (n = 4,262), Cosrf (n = 842), MGC (n = 72), LOC (n = 2,503), 12,808 well characterized detected genes remained for analysis.

**Genome wide association analysis**

To test all associations between SNPs and expressions in a reasonable amount of time, a C script calling the GNU Scientific Library (GSL) “TAMU ANOVA” (www.stat.tamu.edu/~ared/ tamuanova/) was written. For all significant associations, results were checked against the R-lm library [56]. When the numbers of homozygotes for the minor allele of a SNP was lower than 30, were checked against the R-lm library [56]. When the numbers of homozygotes for the minor allele of a SNP was lower than 30, they were grouped with heterozygotes. We used a family-wise error rate of 0.05 corrected for the number of tested SNP x expression associations, which corresponds to declare significant any association with a P-value <5.78x10^{-12}. To increase robustness, associations significant by ANOVA were further checked by a Kruskall-Wallis (KW) test and only associations with a P-value <10^{-10} by the KW test were retained. P-values given in the results and in the GHS-Express database are those obtained by ANOVA. For SNPs on chromosome X, associations with gene expression were assessed separately in women and men and the P-values were combined using the Fisher method [57].

**Association of gene expressions with CVD risk factors**

The relationship between gene expression and each CHD risk factor was tested by a linear regression model using R-lm, with gene expression as the dependent variable. Association with age was adjusted for gender while association with other risk factors were adjusted for gender, age and, if specified, BMI. A square root transformation was applied to CRP levels to remove positive skewness. A study-wise statistical significance threshold of 3.9x10^{-7} was used to correct for the number of tests (10 risk factors x12,808 gene expressions). For each expression trait that was associated with a risk factor and also affected by cis eSNPs, we tested the interaction between the risk factor and the best cis eSNP on expression in a regression model.

Global assessment of associations between the monocyte transcriptome and CVD risk factors

The goal of this analysis was to identify subsets of expression traits independently associated with each risk factor. To increase the robustness of the analysis the population was randomly divided into 2 sub-samples of equal size which were used for screening and validation purpose respectively. The screening step was focused on the subsets of expression traits that were associated with each covariate-adjusted risk factor in univariate analysis at P<3.9x10^{-12} (Bonferroni correction for 12,808 expressions). Each risk factor and corresponding subset of expression traits were included as dependent and predictor variables respectively in a forward stepwise regression model to identify expression traits that were independently associated with the risk factor (P<0.01). Gene expressions selected at the screening step were then jointly tested in the validation sample for association with the risk factor by multiple regression analysis. This screening/validation procedure was repeated 250 times and for each risk factor, expression traits associated (P<0.01) with the risk factor in more than 25% of the replicates are reported.

**Power of the SNP-expression association analysis**

Power was calculated using the program Quanto (http://hydra.usc.edu/GxE/). Assuming a quantitative expression trait with mean 0 and SD 1, a sample size of 1,490 subjects, a type I error of 5.78x10^{-12} and an additive allele effect, the study had a 82% power to detect the effect of a SNP explaining 4% of gene expression.

**Functional classification of genes**

An ontology analysis was performed using the Panther database (http://www.pantherdb.org/). Lists or sublists of genes involved in associations with eSNPs or risk factors were compared to the background list of the 12,808 genes. The P-value calculated by the binomial statistic and Bonferroni-corrected was used.

**Quality checking and exclusion of outliers**

Population stratification and quality of genotypes and expression data were tested extensively and outliers were excluded on the basis of multidimensional scaling analysis (see Methods S2)

**GHS_Express**

A downloadable SQL database compiling the results of the various associations tested is available online (http://genecanvas.ecgene.net/uploads/ForReview/), see also Methods S1. This database can be used to test specific hypotheses.

**Supporting Information**

**Table S1** Characterization of polymorphic probes in eQTLs. Found at: doi:10.1371/journal.pone.0010693.s001 (0.33 MB DOC)

**Table S2** Loci identified in GWAS of BMI and BP - associations of lead/tag SNPs with phenotypes and expressions, and of expressions with phenotype in GHS. Found at: doi:10.1371/journal.pone.0010693.s002 (0.07 MB DOC)

**Table S3** Sets of gene expressions robustly and independently associated with each risk factor in the validation samples. Found at: doi:10.1371/journal.pone.0010693.s003 (0.21 MB DOC)

**Methods S1** GHS-Express database.
Methods S2  Data quality checking.

Found at: doi:10.1371/journal.pone.0010693.s005  (0.21 MB DOC)

File S1  GHS - Characteristics of cis eQTLs.

Found at: doi:10.1371/journal.pone.0010693.s006  (0.40 MB XLS)

File S2  Comparison cis eQTL in GHS and the study of Stranger et al.

Found at: doi:10.1371/journal.pone.0010693.s007  (0.10 MB XLS)

File S3  Comparison cis eQTL in GHS and the study of Dixon et al.

Found at: doi:10.1371/journal.pone.0010693.s008  (0.14 MB XLS)

File S4  Comparison cis eQTL in GHS and the study of Schadt et al.

Found at: doi:10.1371/journal.pone.0010693.s009  (0.28 MB XLS)

File S5  Comparison cis eQTL in GHS and the study of Goring et al.

Found at: doi:10.1371/journal.pone.0010693.s010  (0.83 MB XLS)

File S6  GHS - Expression traits associated with risk factors.

Found at: doi:10.1371/journal.pone.0010693.s011  (0.29 MB XLS)

File S7  GHS - Interaction cis eSNPs with risk factors.

Found at: doi:10.1371/journal.pone.0010693.s012  (0.12 MB XLS)

Acknowledgments

We thank John Todd and Chris Wallace for very helpful comments on our manuscript. We acknowledge Carolin Neukirch, Fatma Karaman, Judita Bühr and Stefanie Müller for technical assistance, Andreas Weith for help during technical performance of GWV and GWE experiments and Alexandru Munteanu for assistance in informatics.

Author Contributions

Conceived and designed the experiments: TZ KJH JL NH TM SB FC. Performed the experiments: TZ PSW CRS RBS DM WR CP CP SB. Analyzed the data: SS MR AS NG MG ME EN DT AZ LT FC. Contributed reagents/materials/analysis tools: TZ PSW KJL HR ME CRS RBS EM WR AZ. Wrote the paper: TZ LT SB FC.

References

1. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–24.
2. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. Nat Genet 39: 1202–7.
3. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. (2006) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 4: e107.
4. Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, et al. (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet 39: 1208–16.
5. Emilsson V, Thorleifsson G, Zhang B, Leonardsson AS, Zink F, et al. (2008) Genetics of gene expression and its effect on disease. Nature 452: 423–8.
6. Idaghdour Y, Cizka W, Shianna KV, Lee SH, Vischer PM, et al. (2010) Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat Genet 42: 62–7.
7. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet 10: 184–94.
8. Nica AC, Dermitzakis ET (2008) Using gene expression to investigate the genetic basis of complex disorders. Hum Mol Genet 17: R129–134.
9. Cheung VG, Spielman RS (2009) Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat Rev Genet 10: 595–604.
10. Gilad Y, Riklin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet 24: 408–415.
11. Crijns AR, Barra D, Milagro FI, Goyenechea E, Larraete E, et al. (2008) Differential expression of oxidative stress and inflammation related genes in peripheral blood mononuclear cells in response to a low-calorie diet: a nutrigenomics study. OMICS 12: 251–261.
12. Bittner M, Misis S, Funte K (2007) Gene expression profiles of T lymphocytes are sensitive to the influence of heavy smoking: A pilot study. Immunogenetics 59: 37–43.
13. Capell F, Klímková E, Viguerie N, Roussel B, Voltóvá M, et al. (2009) Macrophages and adipocytes in human obesity: adipose tissue gene expression and insulin sensitivity during caloric restriction and weight stabilization. Diabetes 58: 1558–1567.
14. de Magalhães JP, Caruso J, Church GM (2009) Meta-analysis of age-related gene expression and regulation of sexually dimorphic genes in mice. Genome Res 19: 25–41.
15. Tan Q, Zhao J, Li S, Christiansen I, Kruse TA, et al. (2008) Differential and correlation analyses of microarray gene expression data in the CEPH Utah families. Genomics 92: 94–100.
16. Yang X, Schadt EE, Wang S, Wang H, Arnold AP, et al. (2006) Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res 16: 995–1004.
17. Ellgren H, Parsch J (2007) The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet 8: 689–698.
18. Weber C, Zernecke A, Libby P (2008) The multifaceted contributions of leukocyte subsets to atherosclerotic lesions from mouse models. Nat Rev Immunol 8: 802–15.
19. Sigurdsson S, Nordmark G, Garnier S, Grundberg E, Kwan T, et al. (2008) A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. Hum Mol Genet 17: 2688–2696.
20. Handmoumuthi L, Ramagopalan SV, Ebers GC, Knight JC (2010) Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. Genes Immun 11: 99–112.
21. Aulchenko YS, Ripatti S, Lindquist I, Boomsma D, Heid IM, et al. (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 41: 47–55.
22. Newton-Clark H, Johnson T, Gateva V, Tobin MD, Bochud M, et al. (2009) Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 41: 666–76.
23. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, et al. (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 41: 18–24.
24. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musumura K, et al. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 41: 56–65.
25. Linsel-Nitschke P, Heeren J, Aherrahrou Z, Bruse P, Gieger C, et al. (2009) Genetic variation at chromosome 1p13.3 affects sortilin mRNA expression, cellular LDL-uptake and serum LDL levels which translates to the risk of coronary artery disease. Atherosclerosis 208: 183–9.
26. Samani NJ, Braund PS, Erdmann J, Götz A, Tomaszewski M, et al. (2008) The novel genetic variant predisposing to coronary artery disease in the region of the PSK1 and CELSR2 genes on chromosome 1 associates with serum cholesterol. J Mol Med 86: 1233–1241.
27. Trégouët D, König IR, Erdmann J, Munteanu A, Braund PS, et al. (2009) Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. Nat Genet 41: 280–282.
28. Kathiresan S, Voight BF, Purcell S, Musumura K, Ardliss D, et al. (2009) Novel genetic variant predisposing to coronary artery disease in the region of the PSK1 and CELSR2 genes on chromosome 1 associates with serum cholesterol. J Mol Med 86: 1233–1241.
29. Kathiresan S, Voight BF, Purcell S, Musumura K, Ardliss D, et al. (2009) Novel genetic variant predisposing to coronary artery disease in the region of the PSK1 and CELSR2 genes on chromosome 1 associates with serum cholesterol. J Mol Med 86: 1233–1241.
30. Visel A, Zhou Y, May D, Alizad V, Gong E, et al. (2010) Targeted deletion of the CD115+ CD135+ common macrophage/DC precursors and the role of CX3CR1 in their response to inflammation. J Exp Med 206: 595–606.
31. Gordon S, Taylor PR (2005) Monocyte and macrophage heterogeneity. Nat Rev Immunol 5: 953–64.
34. Zhou T, Chen Y, Hao L, Zhang Y (2006) DC-SIGN and immunoregulation. Cell Mol Immunol 3: 279–83.
35. Collot-Teixeira S, Martin J, McDermott-Roe C, Poston R, McGregor JL (2007) CD36 and macrophages in atherosclerosis. Cardiovasc Res 75: 468–77.
36. Zudaire E, Portal-Núñez S, Caritusa F (2006) The central role of adrenomedullin in host defense. J Leukoc Biol 80: 237–44.
37. Chan Y, Chiang M, Tsai Y, Su S, Chen M, et al. (2009) Absence of the Transcriptional Repressor Blimp-1 in Hematopoietic Lineages Reveals Its Role in Dendritic Cell Homeostatic Development and Function. J Immunol 183: 7039–7046.
38. He JQ, Wiersmann C, van Lookeren Campagne M (2008) A role of macrophage complement receptor CR1g in immune clearance and inflammation. Mol Immunol 45: 4041–4047.
39. Auffray C, Sieweke MH, Geissmann F (2009) Blood Monocytes: Development, Heterogeneity, and Relationship with Dendritic Cells. Annu Rev Immunol 27: 669–692.
40. Swirski FK, Weissleder R, Pittet MJ (2009) Heterogeneous In Vivo Behavior of Monocyte Subsets in Atherosclerosis. Arterioscler Thromb Vasc Biol 29: 1424–32.
41. Combadie`re C, Potteaux S, Gao J, Esposito B, Casanova S, et al. (2003) Decreased atherosclerotic lesion formation in CX3CR1/apolipoprotein E double knockout mice. Circulation 107: 1009–16.
42. Oram JF, Vaughan AM (2006) ATP-Binding cassette cholesterol transporters and cardiovascular disease. Circ Res 99: 1031–43.
43. Schmutz G, Langmann T, Heimerl S (2001) Role of ABCG1 and other ABCG family members in lipid metabolism. J Lipid Res 42: 1513–1520.
44. Zelcer N, Hong G, Boyadjian R, Tontonoz P (2009) LXR regulates cholesterol uptake through Idol-dependent ubiquitination of the LDL receptor. Science 325: 100–104.
45. Lindholm D, Bornhauser BC, Korhonen I (2009) Mylpi makes an Idol turn into regulation of LDL receptor. Cell Mol Life Sci 66: 3399–3402.
46. Flowers MT, Ntabi JM (2006) Role of stearoyl-coenzyme A desaturase in regulating lipid metabolism. Curr Opin Lipidol 19: 248–256.
47. Watanabe M, Layne MD, Haich C, Maemura K, Gray S, et al. (2002) Regulation of smooth muscle cell differentiation by AT-rich interaction domain transcription factors Mrf2alpha and Mrf2beta. Circ Res 91: 382–389.
48. Lasky-Su J, Neale BM, Franke B, Anney RJL, Zhou K, et al. (2008) Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. Am J Med Genet B Neuropsychiatr Genet 147B: 1345–1354.
49. Kimkis C, Martini M, Friederichs J, Rosenberg R, Doll D, et al. (2006) Prognostic significance of downregulated expression of the candidate tumour suppressor gene SASH1 in colon cancer. Br J Cancer 95: 1419–1423.
50. Cox MA, Gomes B, Palmer K, Du K, Wiekowski M, et al. (2005) The pyrimidinergic P2Y6 receptor mediates a novel release of proinflammatory cytokines and chemokines in mononuclear cells stimulated with UDP. Biochem Biophys Res Commun 330: 467–73.
51. Charles PC, Alder BD, Hilliard EG, Schinder JC, Lineberger RE, et al. (2008) Tobacco use induces anti-apoptotic, proliferative patterns of gene expression in circulating leukocytes of Caucasian males. BMC Med Genomics 1: 38.
52. Amброс JA, Barua RS (2004) The pathophysiology of cigarette smoking and cardiovascular disease: an update. J Am Coll Cardiol 43: 1731–1737.
53. Dinas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325: 1246–1250.
54. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215.
55. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 Suppl 1: S96–104.
56. R Development Core Team (2008) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing Available at: http://www.R-project.org.
57. Fisher R (1932) Statistical Methods for Research Workers. London: Oliver and Boyd.