Genome-Wide Association Study for Endothelial Growth Factors

Wolfgang Lieb, MD, MSc; Ming-Huei Chen, PhD; Martin G. Larson, ScD; Radwan Safa, MD; Alexander Teumer, PhD; Sebastian E. Baumeister, PhD; Honghuang Lin, PhD; Holly M. Smith, MS; Manja Koch, PhD; Roberto Lorbeer, PhD; Uwe Völker, PhD; Matthias Nauck, MD; Henry Völzke, MD; Henri Wallaschofski, MD; Douglas B. Sawyer, MD, PhD; Ramachandran S. Vasan, MD

Background—Endothelial growth factors including angiopoietin-2 (Ang-2), its soluble receptor Tie-2 (sTie-2), and hepatocyte growth factor play important roles in angiogenesis, vascular remodeling, local tumor growth, and metastatic potential of various cancers. Circulating levels of these biomarkers have a heritable component (between 13% and 56%), but the underlying genetic variation influencing these biomarker levels is largely unknown.

Methods and Results—We performed a genome-wide association study for circulating Ang-2, sTie-2, and hepatocyte growth factor in 3571 Framingham Heart Study participants and assessed replication of the top hits for Ang-2 and sTie-2 in 3184 participants of the Study of Health in Pomerania. In multivariable-adjusted models, sTie-2 and hepatocyte growth factor concentration were associated with single-nucleotide polymorphisms in the genes encoding the respective biomarkers (top \( P=2.40\times10^{-45} \) [rs2273720] and 3.64\( \times10^{-19} \) [rs7545687], respectively). Likewise, rs2442517 in the MCPH1 gene (in which the Ang-2 gene is embedded) was associated with Ang-2 levels (\( P=5.05\times10^{-8} \) in Framingham Heart Study and 8.39\( \times10^{-5} \) in Study of Health in Pomerania). Furthermore, single-nucleotide polymorphisms in the AB0 gene were associated with sTie-2 (top single-nucleotide polymorphism rs8176693 with \( P=1.84\times10^{-33} \) in Framingham Heart Study; \( P=2.53\times10^{-30} \) in Study of Health in Pomerania) and Ang-2 (rs8176746 with \( P=2.07\times10^{-8} \) in Framingham Heart Study; \( P=0.001 \) in Study of Health in Pomerania) levels on a genome-wide significant level. The top genetic loci were explained between 1.7% (Ang-2) and 11.2% (sTie-2) of the interindividual variation in biomarker levels.

Conclusions—Genetic variation contributes to the interindividual variation in growth factor levels and explains a modest proportion of circulating hepatocyte growth factor, Ang-2, and Tie-2. This may potentially contribute to the familial susceptibility to cancer, a premise that warrants further studies. (Circ Cardiovasc Genet. 2015;8:389-397. DOI: 10.1161/CIRCGENETICS.114.000597.)

Key Words: angiopoietin-2 ■ endothelial growth factors ■ genetics ■ Genome-Wide Association Study ■ hepatocyte growth factor ■ Tie-2 receptor

Clinical Perspective on p 397

Endothelial growth factors play an important role in human physiology. Because of their pivotal effects on angiogenesis and vascular remodeling,1 endothelial growth factors have been implicated in early vascular development,2 the local spread/growth and metastatic potential of various cancers,3,4 and the pathogenesis of chronic conditions, including cardiovascular and metabolic diseases.5

Previous evidence suggests that circulating levels of these endothelial growth factors are heritable. Specifically, 56%, 38%, and 27% of the interindividual variation in soluble Tie-2 (sTie-2), hepatocyte growth factor (HGF), and angiopoietin-2 (Ang-2) concentrations, respectively, were explained by additive genetic factors.5,7 However, the precise genetic underpinnings of circulating Ang-2, Tie-2, and HGF levels are largely unknown. A previous family-based analysis identified significant linkage of sTie-2 levels to a
relatively broad locus on chromosome 9, including the Tie-2 gene (logarithm of the odds score, 8.31).\(^7\) In the present analyses, we report the results of a genome-wide association study for serum levels of Ang-2, sTie-2, and HGF (discovery sample, n=3571; independent replication sample for Ang-2 and sTie-2, n=3184). Given the moderate to substantial heritability estimates,\(^8,9\) we hypothesized that circulating concentrations of all the 3 biomarkers will be associated with common genetic variants.

### Methods

#### Discovery Sample (Framingham Heart Study)

The initial genome-wide analyses were performed in participants of the Third Generation cohort of the Framingham Heart Study (FHS).\(^4\) At the baseline examination cycle (2002–2005), participants were comprehensively phenotyped, including assessment of standard cardiovascular disease risk factors, a physical examination, and a standardized interview by an FHS physician. The sample sizes available for genome-wide analyses in the discovery sample varied slightly between biomarkers depending on the availability of covariates: HGF, n=3571; Ang-2, n=3574; and sTie-2, n=3574. Written informed consent (including consent to genetic analyses) was provided by all participants and the Institutional Review Board at the Boston University Medical Center approved the study protocol.

#### Replication Sample (Study of Health in Pomerania)

Replication of single-nucleotide polymorphisms (SNPs) that reached genome-wide significance (\(P<5\times10^{-8}\)) for association with Ang-2 or sTie-2 levels in the discovery sample (FHS sample) was assessed in participants of the second examination cycle (2002–2006) of the Study of Health in Pomerania (SHIP).\(^1\) SHIP is a community-based sample in Northeast Germany. The study design has been described in detail elsewhere.\(^9\) In essence, SHIP represents a random cluster sample, representative of the population of West Pomerania, located in Northeast Germany. The response rate of the SHIP-1 examination cycle was 83.6%.\(^9\) The SHIP study was approved by the Ethics Committee of the University of Greifswald. Previous genetic analyses had shown that the SHIP sample is genetically homogeneous and provided no evidence for a genetic drift.\(^10,11\) Furthermore, the design effects (because of cluster sampling) on the \(P\) values caused by deviation from simple random sampling were negligible. Therefore, the cluster sampling was not taken into account in our genetic analyses.

#### Biomarker Measurement

In the FHS sample, blood was drawn in the early morning (usually between 7:00 and 9:00 AM) after an overnight fast, immediately centrifuged, and stored at \(-80^\circ\text{C}\) until biomarkers were assayed. Ang-2, sTie-2 and HGF were measured using commercial assays (R&D Inc), as described previously.\(^5,7\) The average interassay coefficients of variation for all the 3 biomarkers were low: 5.7% for Ang-2, 3.2% for Tie-2, and 1.6% for HGF.\(^6,7\) In SHIP, Ang-2 and Tie-2 were measured likewise using commercially available R&D assays. The minimum detectable concentration was 1.20 pg/mL and 0.001 ng/mL for Ang-2 and for Tie-2, respectively. The interassay coefficients of variation in SHIP were 10.6% and 13.3% for Ang-2, and 5.6% and 6.7% for Tie-2, for low and high biomarker concentrations, respectively.\(^1\)

#### Genotyping, Imputation, and Controlling for Population Stratification

FHS participants were genotyped with the Affymetrix Human Mapping 500K Array Set and the 50K Human Gene Focused Panel. The mean call rate was 98%.\(^1,3\) SNPs with a call rate <95% or deviation from Hardy–Weinberg equilibrium (\(P<10^{-6}\)) were excluded. As additional quality control measures in FHS and SHIP, the reported sex in the database had to agree with the called sex in the genetic data set; identity by descent estimations were performed to identify and exclude duplicated individuals, and individuals with a high genome-wide heterozygosity rate (beyond 5 SDs of the mean) were removed before analyses.\(^12,13\) Genotypes were imputed to the 
\(n=2.5\) million SNP of the HapMap CEU panel (release 22; build 36; http://hapmap.ncbi.nlm.nih.gov/) using the Markov Chain Haplotyping (MACH) algorithm.\(^16\) Quantile–quantile plots and \(P\) estimates were obtained to assess whether the observed distribution of \(P\) values fits the \(P\)-value distribution expected under the null hypothesis of no association, except at the extreme tail. Principal component analyses were performed using EIGENSTRAT\(^17\) to account for potential population substructures within the FHS data set, but no admixture was observed in the FHS cohort.

SHIP participants were genotyped with the Affymetrix Human SNP Array 6.0 using the Birdseed2 clustering algorithm.\(^11\) Pair-wise linkage disequilibrium between SNPs was assessed using the SNP annotation and proxy search (SNAP) tool.\(^18\) Details of the exact location of SNPs within or between genes were assessed using the UCSC genome browser (http://genome.ucsc.edu/).

### Statistical Analyses

Because of their skewed distributions, all biomarkers were natural logarithmically transformed before performing genetic analyses. Plots of untransformed and transformed concentrations of HGF, Ang-2, and Tie-2 in the FHS sample (Figures IA–IIIA in the Data Supplement) and in the SHIP sample (Figures IB–IIB in the Data Supplement) are provided in Figures I–III in the Data Supplement. Assuming an additive genetic model with 1 degree of freedom, we related each of the 2.5 million SNPs in the FHS sample to each circulating biomarker (HGF, Ang-2, and sTie-2), using a linear mixed-effects model that accounts for the family relatedness (R package Genome-Wide Association analyses with Family [GWAF]).\(^19\) Specifically, we accounted for familial correlations with a linear mixed-effects model that uses relationship coefficient matrix as within pedigree correlation matrix. The relationship coefficient (which is twice the kinship coefficient) was calculated based on available pedigree information and structure. Multivariable analyses were adjusted for age, sex, and known clinical correlates of each biomarker, as identified in previous analyses in the FHS data set.\(^5,7\) These additional covariates were as follows: for Ang-2: smoking, total cholesterol, diastolic and systolic blood pressure, antihypertensive treatment, and diabetes mellitus; for Tie-2: diabetes mellitus, body mass index, estimated glomerular filtration rate, alcohol consumption, and triglycerides; and for HGF: diastolic blood pressure, antihypertensive treatment, high-density lipoprotein cholesterol, triglycerides, smoking, body mass index, and diabetes mellitus. A \(P<5\times10^{-8}\) was considered genome-wide significant. A top SNP (rs2442517) of the Ang-2 genome-wide association study with \(P=5.05\times10^{-8}\) was also listed as the \(P\) value was close to the genome-wide significance threshold. For each biomarker, we graphically plotted each SNP’s \(P\) value versus its physical position (Manhattan plot).

The top SNP of each locus associated with Ang-2 and Tie-2 levels in the FHS sample in a genome-wide significant fashion was tested for replication in the SHIP sample, likewise using a linear regression model assuming an additive genetic model and adjusting for the same set of covariates as described above. We did not assess replication of the top SNPs related to HGF concentrations because HGF levels were not available in SHIP.

### Pathway Analyses

Pathway analyses were conducted for Ang-2 and sTie-2 combined and separately for HGF using the program Ingenuity Pathway Analysis software package (Ingenuity Systems, Redwood, CA). A SNP score was assigned to each genetic variant. This SNP score was equivalent to the lower \(P\) value of the 2 traits (Ang-2 and Tie-2, respectively)
examined. For the pathway analysis for HGF, the SNP score was equivalent to the $P$ value for association with HGF. On the basis of the human reference genome (NCBI Build 36, 2006), the location of all SNPs relative to RefSeq genes were obtained from UCSC Genome Browser (http://genome.ucsc.edu/). A gene score was defined as the most significant variant that was located within 110 kb upstream and 40 kb downstream of the gene’s most extreme transcript boundaries. If a SNP could be mapped to multiple genes, all these genes received the same gene score (the $P$ value of this SNP) if the respective SNP had the lowest $P$ value in the region. Of the $\approx 23,000$ genes evaluated, 234 genes and 294 genes reached a score $<1.0 \times 10^{-4}$ for HGF and for Ang-2/sTie-2, respectively. These genes were then used for pathway analyses. Enrichment of each of the canonical pathways for HGF- and Ang-2/sTie-2–related SNPs, respectively, was assessed using Fisher exact test.

Statistical analyses were performed in R. The authors had full access to the data and take responsibility for its integrity. M.H.C. performed the statistical analyses for the FHS sample, and A.T. performed the statistical analyses for the SHIP sample. The pathway analyses were performed by H.L. All authors have read and agreed to the article as written.

### Results

Clinical and biochemical characteristics of the discovery and the replication sample are provided in Table 1. In the discovery sample, the genomic inflation factor did not indicate systematic inflation for each of the biomarkers (between 1.01 and 1.02; Figure IV in the Data Supplement). The top

### Table 1. Clinical and Biochemical Characteristics of the Discovery and Replication Samples

<table>
<thead>
<tr>
<th>Variables</th>
<th>Discovery Sample (FHS Third Generation)</th>
<th>Replication Sample (SHIP-1)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women (n=1904)</td>
<td>Men (n=1670)</td>
</tr>
<tr>
<td><strong>Clinical features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, y</td>
<td>40±9</td>
<td>40±9</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>113±14</td>
<td>121±12</td>
</tr>
<tr>
<td>Diastolic blood pressure, mm Hg</td>
<td>73±9</td>
<td>78±9</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>26±6</td>
<td>28±5</td>
</tr>
<tr>
<td>Waist circumference, inches</td>
<td>35±6</td>
<td>39±5</td>
</tr>
<tr>
<td>Smoking, %</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Diabetes mellitus, %</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Biochemical features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cholesterol, mg/dL</td>
<td>185±34</td>
<td>193±37</td>
</tr>
<tr>
<td>HDL cholesterol, mg/dL</td>
<td>61±16</td>
<td>47±12</td>
</tr>
<tr>
<td>Triglycerides, mg/dL</td>
<td>97±63</td>
<td>134±106</td>
</tr>
<tr>
<td>Alcohol consumption†</td>
<td>5.8±7.9</td>
<td>13.8±106</td>
</tr>
<tr>
<td>eGFR, mL/min per 1.73 m²</td>
<td>99±19</td>
<td>100±17</td>
</tr>
<tr>
<td><strong>Endothelial growth factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGF, pg/mL</td>
<td>825 (697, 974)</td>
<td>814 (702, 967)</td>
</tr>
<tr>
<td>Angiopoietin-2, ng/mL</td>
<td>2.02 (1.51, 2.71)</td>
<td>1.70 (1.33, 2.23)</td>
</tr>
<tr>
<td>Soluble Tie-2, ng/mL</td>
<td>14.3 (12.0, 17.3)</td>
<td>15.0 (12.7, 18.2)</td>
</tr>
</tbody>
</table>

Data are mean±SD; except for HGF, Angiopoietin-2, and soluble Tie-2 where median (quartile 1 and quartile 3) are provided. eGFR indicates estimated glomerular filtration rate; FHS, Framingham Heart Study; HDL, high-density lipoprotein; HGF, hepatocyte growth factor; and SHIP, Study of Health in Pomerania.

*Genetic data were missing in 16 participants.

†In ounces per month in FHS sample, in g/d in the SHIP-1 sample.

### Table 2. Genome-Wide Significant Loci for Circulating Concentrations of HGF, sTie-2, and Ang-2

| Biomarker | rsID | Chr | n | Coded/Noncoded | MAF | $|i| (SE)$ | $P$ Value | Closest Gene | n | MAF | $|i| (SE)$ | $P$ Value |
|-----------|------|-----|---|---------------|-----|--------|---------|------------|---|-----|--------|---------|
| HGF       | rs5745687 | 7   | 3571 | T/C | 0.077 | -0.099 (0.011) | $3.64 \times 10^{-13}$ | HGF | 3188 | 0.096 | 0.061 (0.019) | 0.001 |
| Ang-2     | rs8176746 | 9   | 3574 | T/G | 0.077 | 0.111 (0.020) | $2.07 \times 10^{-8}$ | ABO | 3188 | 0.461 | 0.043 (0.011) | $8.39 \times 10^{-5}$ |
|           | rs2442517 | 8   | 3574 | G/A | 0.474 | 0.059 (0.011) | $5.05 \times 10^{-4}$ | MCPH1 | 3188 | 0.056 | 0.237 (0.018) | $2.70 \times 10^{-3}$ |
| sTie-2    | rs2273720 | 9   | 3574 | C/A | 0.056 | 0.265 (0.016) | $2.40 \times 10^{-5}$ | Tie-2 | 3184 | 0.096 | 0.157 (0.014) | $2.53 \times 10^{-3}$ |
|           | rs8176693 | 9   | 3574 | T/C | 0.077 | 0.155 (0.013) | $1.84 \times 10^{-3}$ | ABO | 3184 | 0.096 | 0.157 (0.014) | $2.53 \times 10^{-3}$ |

The most significant single-nucleotide polymorphisms for each locus and biomarker are shown. Ang-2 indicates Angiopoietin-2; Chr, chromosome; FHS, Framingham Heart Study; HGF, hepatocyte growth factor; MAF, minor allele frequency; SHIP, Study of Health in Pomerania; and sTie-2, soluble Tie-2.
genome-wide significant (P ≤ 5×10⁻⁸) hits for Ang-2, sTie-2, and HGF were listed in Table 2. Figure 1 displays the P values for ≈ 2.5 million genetic variants associated with HGF, Ang-2, and sTie-2, respectively.

Genetic Loci Associated With Circulating HGF Concentrations

In the FHS sample, 29 SNPs within or close to the HGF gene on chromosome 7 were associated with circulating HGF levels on a genome-wide significant level (P ≤ 5×10⁻⁸), most of these SNPs were in moderate linkage disequilibrium with the top SNP (rs5745687; Table 2; Figure 2). Overall, 24 of these 29 SNPs were genotyped in the FHS sample, and 5 SNPs were imputed (including rs5745687). A regional plot is shown in Figure 2. The top SNP in the HGF gene explained 2.12% of the variation in circulating HGF levels. Mean HGF levels stratified by rs5745687 genotypes are provided in Figure V in the Data Supplement.

Genetic Loci Associated With Circulating Ang-2 and sTie-2 Concentrations

A total of 5 SNPs in the ABO gene on chromosome 9 were associated with circulating Ang-2 concentrations on a genome-wide significant level (P ≤ 5×10⁻⁸); 4 of these SNPs were genotyped in the FHS sample. The top SNP was imputed in the FHS and SHIP samples; rs2442517 in the MCPH1 gene on chromosome 8 was close to the genome-wide significance level (P = 5.05×10⁻⁸). This SNP was genotyped in the FHS sample and imputed in the SHIP sample. Tie-2 levels stratified by rs2273720 genotypes (Figure VIIA in the Data Supplement) and stratified by rs8176693 genotypes (Figure VIIB in the Data Supplement) in the FHS and the SHIP samples are provided in Figure VII in the Data Supplement. Overall, 17 SNPs within or close to the Tie-2 gene and 56 SNPs within or close to the ABO gene were associated with sTie-2 levels on a genome-wide significant level (P ≤ 5×10⁻⁸; Table 2). Regional plots of the Tie-2 and the ABO loci are displayed in Figure 4A and 4B, respectively. The most significant SNPs from each significant locus (rs8176693 and rs2273720) combined explained 11.2% and 8.9% of the interindividual variation in circulating Tie-2 levels in the discovery sample and in the replication sample, respectively. SNP rs2273720 was genotyped in the FHS and SHIP samples, SNP rs8176693 was genotyped in FHS and imputed in the SHIP sample. Tie-2 levels stratified by rs2273720 genotypes (Figure VIIA in the Data Supplement) and stratified by rs8176693 genotypes (Figure VIIB in the Data Supplement) in the FHS and the SHIP samples are provided in Figure VII in the Data Supplement.

![Figure 1](http://circgenetics.ahajournals.org/)

**Figure 1.** Manhattan plot displaying the association of ≈2.5 million single-nucleotide polymorphisms (SNPs) with circulating concentrations of hepatocyte growth factor (HGF), angiopoietin-2, and soluble Tie-2 (sTie-2) in the Framingham Heart Study sample. Each dot represents a SNP. On the x axis, the chromosomal location of the genetic variant is shown (with the results being plotted from the P-terminal end). On the y axis, the −log₁₀ (P value) is shown. The gray horizontal line indicates the threshold of genome-wide significance (5×10⁻⁸).

![Figure 2](http://circgenetics.ahajournals.org/)

**Figure 2.** Regional plot displaying the association of single-nucleotide polymorphisms (SNPs) at the hepatocyte growth factor (HGF) gene locus with circulating HGF levels. The colors represent the linkage disequilibrium (expressed as r²) between neighboring SNPs. The gray horizontal line indicates the threshold of genome-wide significance (5×10⁻⁸).
Lieb et al
GWAS for Endothelial Growth Factors
393

(rs8176746) were in perfect linkage disequilibrium ($r^2=1$), underscoring that Ang-2 and Tie-2 levels were associated with the same genetic locus.

Pathway Analyses
For Ang-2 and sTie-2, no significantly enriched canonical pathway ($P<1\times10^{-3}$) could be identified.

By contrast, several pathways that were over-represented with HGF-related genes could be observed (Table I in the Data Supplement), including the mammalian target of rapamycin (mTOR) pathway and the amyotrophic lateral sclerosis signaling pathway.

Discussion
Principal Observations
We conducted a genome-wide association study for 3 endothelial growth factors (Ang-2, sTie-2, and HGF) in the FHS and assessed replication of the top SNPs associated with Ang-2 and sTie-2 in the SHIP study, a population-based sample from northeastern Germany. Our main findings were 4-fold. First, all the 3 biomarker levels were associated with common genetic variants within or close to the genes encoding the individual biomarkers. Second, Ang-2 and sTie-2 levels were also associated with genetic variants at the $ABO$ locus. Third, the top genetic loci explained a modest ($\approx 2\%$; HGF and Ang-2) to moderate (11%; sTie-2) proportion of the interindividual variation in biomarker levels. Fourth, we identified several pathways that were over-represented with HGF-related genes, including the mTOR and the amyotrophic lateral sclerosis signaling pathway.

Genetic Loci Associated With Circulating HGF Concentrations
Previous family-based analyses reported heritability estimates of 38%$^6$ and 48%,$^{20}$ respectively, for HGF levels. In

Figure 3. Regional plots displaying the association of single-nucleotide polymorphism (SNPs) at the microcephalin 1 ($MCPH1$) gene locus (A) and at the $ABO$ locus (B), associated with circulating angiopoietin-2 (Ang-2) levels. The colors represent the linkage disequilibrium (expressed as $r^2$) between neighboring SNPs. $ANGPT2$ indicates Ang-2. The gray horizontal line indicates the threshold of genome-wide significance ($5\times10^{-8}$).

Figure 4. Regional plots displaying the association of single-nucleotide polymorphisms (SNPs) at the Tie-2 (denoted as TEK) gene locus (A) and the $ABO$ locus (B) with circulating Tie-2 levels. The colors represent the linkage disequilibrium (expressed as $r^2$) between neighboring SNPs. The gray horizontal line indicates the threshold of genome-wide significance ($5\times10^{-8}$).
small clinical samples, few genetic variants in the \textit{HGF} gene have been tested for association with HGF levels revealing inconsistent results. In 78 control participants from the Cardiovascular Health Study, SNP rs3735520, which is located in the promoter region of the \textit{HGF} gene, was statistically significantly associated with HGF levels, whereas SNP rs17501108 (likewise located in the promoter) was not.\textsuperscript{21} However, SNP rs3735520 was not associated with HGF levels in 133 patients with systemic sclerosis,\textsuperscript{22} although experimental data indicated that this variant was related to transcriptional activity.\textsuperscript{23} In our data set, both SNP rs3735520 ($P=0.12$) and SNP rs17501108 ($P=0.39$) were not associated with HGF concentrations. To date, there has been no genome-wide analysis about the genetic determinants of HGF levels. In our genome-wide approach, only SNPs in the \textit{HGF} gene reached genome-wide significance for this biomarker. The top SNP (rs5745687) is located in exon 823 and explained $\approx 2.12\%$ of the interindividual variation in HGF levels. According to the pfam database (http://pfam.xfam.org/protein/P14210),\textsuperscript{24} SNP rs5745687 leads to the replacement of a negatively charged glutamic acid by a positively charged lysine at position 304. This missense mutation is located just before the third kringle domain of the HGF protein. Such kringle domains are important for the interaction with other proteins.\textsuperscript{24,25} The frequency of the minor allele of rs5745687 is between 0\% and 10\% in different ethnic subgroups of the 1000 Genomes project\textsuperscript{26} (Table II in the Data Supplement). We did not assess replication of the top SNPs related to HGF concentrations because HGF levels were not available in SHIP.

**Genetic Loci Associated With Circulating Tie-2 and Ang-2 Concentrations**

The statistically strongest association signal in our analyses was observed for SNPs in the \textit{Tie-2} gene with circulating Tie-2 concentrations. The top SNP of the \textit{Tie-2} locus (rs2273720; located in intron 16) explained $\approx 7\%$ of the variability in \textit{sTie-2} levels in the FHS sample and $\approx 5\%$ in the SHIP replication sample. These findings are in good agreement with a recent linkage analysis (displaying strong linkage of \textit{sTie-2} levels to a locus on chromosome 9, including the \textit{Tie-2} gene)\textsuperscript{27} and suggest that \textit{sTie-2} is moderately influenced by genetic variation in the \textit{Tie-2} gene. Furthermore, the \textit{ABO} locus was associated with \textit{sTie-2} on a genome-wide scale (please see below for expanded discussion about the \textit{ABO} locus).

Circulating levels of Ang-2 were associated with rs2442517, an intronic variant in the \textit{microcephalin 1} (\textit{MCPH1}) gene. Interestingly, the \textit{Ang-2} gene is nested within the \textit{MCPH1} gene (please see regional plot in Figure 3A) and the most probable explanation for this association signal is that \textit{Ang-2} levels are in part determined by genetic variation in the \textit{Ang-2} gene. However, the correlation between SNP rs2442517 and variants in the \textit{Ang-2} gene is only modest (please see linkage disequilibrium structure in Figure 3A), and genetic variants in the \textit{Ang-2} gene itself did not reach genome-wide significance for Ang-2 levels. Our data add to smaller genetic-epidemiological studies relating SNPs within the \textit{Ang-2} gene to circulating Ang-2 concentrations. SNP rs3739391 (located in the 5\' untranslated region; weakly correlated with rs2442517; $r^2=0.04$) was associated with Ang-2 levels in 360 stroke patients, after adjusting for established vascular risk factors. However, in the same sample, 2 other genetic variants (rs2515507 [located in the promoter] and rs3739390 [5\' untranslated region]) displayed no association with Ang-2 levels.\textsuperscript{27} In our data set, none of rs3739391 ($P=0.10$), rs2515507 ($P=0.35$), and rs3739390 ($P=0.17$) displayed statistically significant evidence for association with Ang-2 levels. In another study, SNP rs1868554 (also in weak linkage disequilibrium with rs2442517; $r^2=0.005$) was not associated with plasma Ang-2 levels, but was related to plasma variation in isoforms of Ang-2.\textsuperscript{28} Interestingly, this variant displayed some evidence of association with Ang-2 levels in our data set ($P=6.5\times10^{-6}$), although the $P$ value does not meet the threshold for genome-wide significance.

In addition to the \textit{MCPH1} locus, also the \textit{ABO} locus was significantly associated with Ang-2 levels in our data set, as detailed in the next paragraph.

**Association of the \textit{ABO} Locus With Circulating Ang-2 and Tie-2 Concentrations**

Several SNPs in the \textit{ABO} gene were associated with \textit{sTie-2} and Ang-2 levels and the top SNPs of the \textit{ABO} locus associated with \textit{sTie-2} (rs8176693) and Ang-2 (rs8176746) were perfectly correlated ($r^2=1$). Thus, Ang-2 and \textit{sTie-2} add to a growing number of biomarkers that are associated with the \textit{ABO} locus, including von Willebrand factor,\textsuperscript{29,30} soluble intercellular adhesion molecule-1 (ICAM-1),\textsuperscript{31,32} E-selectin,\textsuperscript{33} P-selectin,\textsuperscript{31} tumor necrosis factors $\alpha$,\textsuperscript{34} and Factor VIII.\textsuperscript{35} This suggests that there might be common mechanisms by which the \textit{ABO} gene product modulates circulating biomarker concentrations. The genetic architecture of the \textit{ABO} locus determines the substrate specificity and activity of a glycosyltransferase, which transfers specific carbohydrates to the H antigen, and thereby determines the \textit{ABO} blood group (OMIM: 110300).\textsuperscript{35,36,37}

There are several potential explanations for the observed association between the \textit{ABO} locus and circulating Ang-2 and \textit{sTie-2} levels. Given that the \textit{Tie-2} receptor contains 4 potential N-glycosylation sites,\textsuperscript{37} the \textit{ABO} gene product might modify the \textit{Tie-2} receptor, and thereby affect its binding activity, its cleavage from the endothelial cells, or its renal clearance. These mechanisms have been suggested to explain the association of the \textit{ABO} locus with other biomarkers, including ICAM-1,\textsuperscript{31,38} P-selectin,\textsuperscript{31,38} E-selectin,\textsuperscript{31} and von Willebrand factor.\textsuperscript{31,38}

In a similar fashion, it is possible that glycosylation of the Ang-2 protein affects its binding ability to the endothelial or \textit{sTie-2} receptor, and thereby influences circulating Ang-2 serum levels. Such glycosylation (by the \textit{ABO} gene product) might also modify the Ang-2 clearance rate from the blood stream.\textsuperscript{31}

In addition to the association with circulating biomarkers mentioned above, variation in the \textit{ABO} gene has been associated with different clinical disease traits, including gastric\textsuperscript{39} and pancreatic cancer,\textsuperscript{36} duodenal ulcer,\textsuperscript{40} as well as myocardial infarction\textsuperscript{41} and stroke,\textsuperscript{42} and hematologic traits.\textsuperscript{43} Many of these disease conditions depend directly or indirectly on
angiogenesis and vascular remodeling—processes that are, in part, governed by endothelial growth factors. Thus, the association of these biomarkers with the ABO locus might be a mediating mechanism for the observed associations between genetic variation in the ABO gene and clinical disease events, a premise that is speculative.

Across the 3 biomarkers (HGF, Ang-2, and sTie-2) analyzed, the genome-wide significantly associated genetic variants explained a modest proportion (between 2% and 11%) of the interindividual variation in biomarker levels. This is comparable with other biomarker-related genome-wide analyses, including those for Factor VII, von Willebrand Factor, and ICAM-1, where the identified genome-wide significant loci explained 7.7%, 12.8%, and 8.4%, respectively, of the interindividual biomarker variation.29,32 Additional genetic approaches are warranted to explain parts of the missing heritability,44 including analyses of rare and low-frequency variants or structural genetic variation44 and analyses of all SNPs in the genome.45 Given the reported associations of the analyzed biomarkers with cardiovascular risk factors6,7 and mortality,12 it is of interest, whether these biomarkers and the identified genetic variants (correlated with biomarker levels) improve the prediction of clinically overt cardiovascular events, including myocardial infarction, heart failure, and stroke.

Pathway Analyses
Pathway analyses identified the mTOR signaling pathway as the most significantly enriched canonical pathway for HGF. Our observations are corroborated by experimental studies, indicating that the mTOR pathway is indeed relevant for modulating HGF expression.46,47 Rapamycin (as inhibitor) and Leucin (as stimulator) exert their influence on HGF expression and secretion through the mTOR pathway.46,47 Furthermore, we revealed a statistically significant over-representation of HGF-related genes in the amyotrophic lateral sclerosis signaling pathway. In this context, experimental data support that HGF might improve survival of motor neurons and the clinical course of amyotrophic lateral sclerosis.48–50

Strengths and Limitations
Both the study samples (FHS and SHIP) are large and provide a thorough and comprehensive characterization of their study participants. To the best of our knowledge, this is the first genome-wide analysis for Ang-2, Tie-2, and HGF. Some limitations merit consideration. Our samples consisted of young to middle-aged Europeans and European Americans. The applicability of our observations to other ethnicities or age groups is unclear. Furthermore, each biomarker was measured only once in each individual, possibly leading to some nondifferential misclassification that would bias our results toward the null hypothesis of no association between genetic variants and biomarker levels. We did not evaluate the effect of rare genetic variants (as could be identified by whole-genome sequencing) on biomarker levels.

Conclusions
We identified common genetic variants associated with circulating HGF, Ang-2, and sTie-2 levels. All the 3 biomarkers were related to SNPs located in or close to the genes encoding the respective biomarkers. Furthermore, genetic variants in the ABO gene were associated with Ang-2 and sTie-2 levels. The association findings for Ang-2 and sTie-2 were successfully replicated in an independent sample (SHIP). The genome-wide significantly associated genetic variants explained a modest proportion (between 2% and 11%) of the interindividual variation in biomarker levels. Furthermore, pathway analyses identified the mTOR and other pathways as being over-represented with HGF-related genes. Given the pivotal role of these biomarkers in angiogenesis, vascular remodeling, local cancer growth, and metastatic potential, as well as their potential effect on cardiovascular and metabolic diseases, our findings suggest that genetically determined endothelial growth factor levels might, in part, contribute to the known familial susceptibility of conditions like cancer or cardiovascular disease, a premise that warrants further investigation.

Sources of Funding
This work was supported by the National Heart, Lung, and Blood Institute’s Framingham Heart Study (contract No. N01-HC-25195; principal investigator, Dr Vasan) and its contract with Affymetrix, Inc for genotyping services (contract No. N02-HL-6-4278), and by grants from the National Heart, Lung, and Blood Institute 2K24HL04334, RO1HL080124, RO1HL077477, and RO1HL093328 (Dr Vasan). SHIP is a part of the Community Medicine Research Network, funded by the Federal State of Mecklenburg-West Pomerania and the University Medicine Greifswald. Genome-wide data have been supported by a joint grant from Siemens Healthcare, Erlangen, Germany, and by the Federal State of Mecklenburg, West Pomerania. Furthermore, this work is part of the research project Greifswald Approach to Individualized Medicine (GANI_MED). The GANI_MED consortium is funded by the German Federal Ministry of Education and Research and, by the Ministry of Cultural Affairs of the German Federal State of Mecklenburg, West Pomerania (03IS2061A).

Disclosures
None.

References


CLINICAL PERSPECTIVE

Endothelial growth factors, including angiopoietin-2 (Ang-2), its soluble receptor Tie-2 (sTie-2), and hepatocyte growth factor (HGF), modulate early vascular development, the local growth and metastatic potential of various cancers and the pathogenesis of cardiovascular and metabolic diseases. These growth factors are heritable traits, but the precise genetic underpinnings of their circulating concentrations are largely unknown. We performed a genome-wide association study for circulating Ang-2, sTie-2, and HGF in 3571 Framingham Heart Study participants and assessed replication of the top hits for Ang-2 and sTie-2 in 3184 participants of the Study of Health in Pomerania. Circulating Tie-2 and HGF concentrations were associated with single-nucleotide polymorphisms in the genes encoding the respective biomarkers. Likewise, rs2442517 in the MCPH1 gene (in which the Ang-2 gene is embedded) was associated with Ang-2 levels. Furthermore, single-nucleotide polymorphisms in the ABO gene were associated with sTie-2 and Ang-2 levels. All the top single-nucleotide polymorphisms associated with Ang-2 and sTie-2 could be successfully replicated in Study of Health in Pomerania. The top single-nucleotide polymorphisms explained between 1.7% (Ang-2) and 11.2% (sTie-2) of the interindividual variation in biomarker levels. In conclusion, genetic variation contributes to the interindividual variation in growth factor levels and explains a modest proportion of circulating biomarker concentrations. This may potentially contribute to the familial susceptibility to cancer, a premise that warrants further studies. Ang-2 and sTie-2 add to a growing number of biomarkers associated with genetic variation at the ABO locus.
Genome-Wide Association Study for Endothelial Growth Factors
Wolfgang Lieb, Ming-Huei Chen, Martin G. Larson, Radwan Safa, Alexander Teumer, Sebastian E. Baumeister, Honghuang Lin, Holly M. Smith, Manja Koch, Roberto Lorbeer, Uwe Völker, Matthias Nauck, Henry Völzke, Henri Wallaschofski, Douglas B. Sawyer and Ramachandran S. Vasan

_Circ Cardiovasc Genet._ 2015;8:389-397; originally published online December 31, 2014;
doi: 10.1161/CIRCGENETICS.114.000597
_Circulation: Cardiovascular Genetics_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/8/2/389
Data Supplement (unedited) at:
http://circgenetics.ahajournals.org/content/suppl/2014/12/31/CIRCGENETICS.114.000597.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation: Cardiovascular Genetics_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation: Cardiovascular Genetics_ is online at:
http://circgenetics.ahajournals.org//subscriptions/
Supplementary Figure 1. Untransformed and transformed HGF concentrations (pg/mL) in the Framingham Heart Study (FHS).
Supplementary Figure 2. Untransformed and transformed Ang-2 concentrations (ng/mL) in the Framingham Heart Study (FHS) sample (Panel A) and in the SHIP (Study of Health in Pomerania) sample (Panel B).

Panel A

Panel B
Supplementary Figure 3. Untransformed and transformed sTie-2 concentrations (ng/mL) in the Framingham Heart Study (FHS) sample (Panel A) and in the SHIP (Study of Health in Pomerania) sample (Panel B).

Panel A

Panel B
Supplementary Figure 4. Quantile-quantile plots comparing the observed against the expected –log10 (p-value) distributions associating common genetic variants to circulating levels of hepatocyte growth factor \((\text{HGF})\), angiopoietin-2 \((\text{Ang-2})\), and Tie-2 \((\text{sTie-2})\), respectively, in the Framingham Heart Study.

Ln, biomarkers were logarithmically transformed (to the base e) prior to genetic analyses.
**Supplementary Figure 5.** Circulating concentrations of hepatocyte growth factor (HGF), stratified by rs5745687 genotype in the Framingham Heart Study (FHS) sample.
Supplementary Figure 6. Angiopoietin-2 (Ang-2) concentrations stratified by rs2442517 genotypes (Panel A) and stratified by rs8176746 genotypes (Panel B) in the Framingham Heart Study (FHS) sample and in the SHIP (Study of Health in Pomerania) sample.

Panel A

Panel B
**Supplementary Figure 7.** Tie-2 concentrations stratified by rs2273720 genotypes (Panel A) and stratified by rs8176693 genotypes (Panel B) in the Framingham Heart Study (FHS) sample and in the SHIP (Study of Health in Pomerania) sample.

**Panel A**

![Panel A graph showing Tie-2 concentrations stratified by rs2273720 genotypes.](image)

**Panel B**

![Panel B graph showing Tie-2 concentrations stratified by rs8176693 genotypes.](image)
**Supplementary Table 1.** Top 5 enriched canonical pathways for hepatocyte growth factor (HGF).

<table>
<thead>
<tr>
<th>Name</th>
<th>p-value</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>mTOR Signaling</td>
<td>1.7E-05</td>
<td>10/198 (0.051)</td>
</tr>
<tr>
<td>Amyotrophic Lateral Sclerosis Signaling</td>
<td>5.2E-05</td>
<td>7/103 (0.068)</td>
</tr>
<tr>
<td>Regulation of eIF4 and p70S6K Signaling</td>
<td>9.7E-05</td>
<td>8/164 (0.049)</td>
</tr>
<tr>
<td>EIF2 Signaling</td>
<td>4.3E-04</td>
<td>8/192 (0.042)</td>
</tr>
<tr>
<td>Small Cell Lung Cancer Signaling</td>
<td>6.4E-04</td>
<td>5/85 (0.059)</td>
</tr>
</tbody>
</table>

mTOR denotes mammalian Target of Rapamycin; eIF2 (Eukaryotic Initiation Factor-2)

**Supplementary Table 2.** Frequency of the minor allele of rs5745687 in different ethnic subgroups in the 1000 Genomes data\(^1\) (http://www.1000genomes.org/data).

<table>
<thead>
<tr>
<th>Population/ Sample</th>
<th>T-Allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU (Utah Residents (CEPH) with Northern and Western European Ancestry)</td>
<td>0.0647</td>
</tr>
<tr>
<td>GBR (British in England and Scotland)</td>
<td>0.0730</td>
</tr>
<tr>
<td>IBS (Iberian population in Spain)</td>
<td>0.1071</td>
</tr>
<tr>
<td>MXL (Mexican Ancestry from Los Angeles USA)</td>
<td>0.0547</td>
</tr>
<tr>
<td>CHB (Han Chinese in Beijing, China)</td>
<td>0.0052</td>
</tr>
</tbody>
</table>

**References**