

Aus der Arbeitsgruppe Bioinformatik des
Max-Delbrück-Centrums für Molekulare Medizin (MDC),
Berlin-Buch, in Kooperation mit der Medizinischen Fakultät
der Charité - Universitätsmedizin Berlin

DISSERTATION

**Einfluss genetischer Varianten auf den
Cholesterinstoffwechsel in einer mitteleuropäischen
Bevölkerungstichprobe**

zur Erlangung des akademischen Grades
Doctor rerum medicarum (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät der Charité -
Universitätsmedizin Berlin

von

Frau Dipl.-Math.(FH) Anja Bauerfeind
aus Aachen

Gutachter: 1. Prof. em. Dr. med. J. G. Reich
2. Prof. Dr. R. Meister
3. Priv.-Doz. Dr. B. Müller-Myhsok

Datum der Disputation: 18. Dezember 2006

Vorbemerkung

Das Wissen um Lipoproteine und Arteriosklerose hat sich gerade in den letzten Jahren dynamisch erweitert. Dazu haben insbesondere zellbiologische und molekularbiologische Untersuchungsmethoden, aber auch epidemiologische Studien wesentlich beigetragen. In der vorliegenden Arbeit „Einfluss genetischer Varianten auf den Cholesterinstoffwechsel in einer mitteleuropäischen Bevölkerungsstichprobe“ behandle ich den komplexen Zusammenhang aus genetisch-epidemiologischer Sicht.

Die Arbeit ist in fünf Kapitel (Einleitung, Methoden, Ergebnisse, Diskussion, Zusammenfassung) aufgeteilt. In Kapitel 1 gehe ich auf die Bedeutung der Fettstoffwechselstörungen, einem Hauptrisikofaktor der Herz-Kreislaufkrankungen ein. Dabei werden das kinetische Modell des Lipoprotein-Stoffwechsels als biologische Grundlage, genetische Einflussfaktoren und die Rolle der Kandidatengene herausgestellt. Daraus ergeben sich in diesem Kapitel eine Reihe von Fragestellungen, die in den folgenden Kapiteln beantwortet bzw. diskutiert werden. Kapitel 2 beginnt mit einer Darstellung der Studienpopulation, Genotypen und Phänotypen. Daran schließt sich eine ausführliche Darstellung der hier entwickelten methodischen Herangehensweise an. Die Analysestrategie ist gleichermaßen ein Ergebnis der Arbeit, wie die Befunde selbst. Statistische Kennzahlen und Verfahren werden nachfolgend eingeführt. Die Analyseergebnisse sind in Form von Tabellen und Abbildungen in Kapitel 3 dargestellt und werden in Kapitel 4 diskutiert. Den Leitfaden der Diskussion bildet der einleitend aufgestellte Fragenkatalog. Den Abschluss bildet Kapitel 5 mit einer Zusammenfassung.

Die Forschungsergebnisse des gesamten interdisziplinären Teams wurden in mehreren Publikationen, in internationalen Zeitschriften und auf Kongressen vorgestellt (Bauerfeind et al., 2006, 2002; Knoblauch et al., 2002, 2004; Nürnberg et al., 2004).

Die Rekrutierung von Probanden und die Phänotypisierung erfolgte durch die Franz-Volhard-Klinik (Prof. Luft und Prof. Schuster), die personenschutzgerechte Erfassung durch die Firma Infogen, Berlin (Prof. Schuster). Genotypisierungen wurden durch Infogen und das MDC (Genome Mapping Center, Prof. Nürnberg, jetzt Köln) realisiert und anteilig vom MedSeq-Forschungsprojekt, der FVK, MDC (Prof. Nürnberg) und Infogen, Berlin finanziert.

Meine Arbeiten umfassen die Erstellung eines Masterfiles der Daten (Familienstruktur, Genotypen, Phänotypen, Plausibilität), die Haplotypisierung der familienstrukturierten Genotypen (gemeinsam mit Dr. Klaus Rohde), die mathematisch-statistische Datenanalyse (Entwicklung von genetischen Modellen der Varianzkomponenten, lineare und logistische Regression, Programmierung der Algorithmen mittels der Statistik-Software SAS, SPlus und R). Die gesamten Auswertungen wurden unter Supervision von Prof. Dr. Jens Reich von mir selbständig durchgeführt. Mein Anteil an den Publikationen besteht in der gesamten statistischen Analyse und Modellierung und in der Formulierung sämtlicher Manuskripte (sowohl bei Erstautorschaft als auch bei Zweitautorschaft zusammen mit Dr. Hans Knoblauch und unter Beratung und Endkorrektur durch Prof. Luft und Prof. Reich). Meine Mitarbeit am vorgelegten Forschungsergebnis wurde und wird in den Jahren 2002-2006 durch ein Doktorandenstipendium unterstützt, das im Rahmen des Deutschen Humangenomforschungsprojektes vergeben wurde. Die Dissertation wurde von mir allein erstellt und das Manuskript von Prof. Reich durchgesehen und mit Änderungsvorschlägen versehen.

Mein ganz besonderer Dank gilt Prof. Jens Reich, der mich ermutigt hat, diese Arbeit zu verfassen, der mich geleitet hat, Mathematik zu benutzen, infrage zu stellen, zu diskutieren und für die Weisheit, Wissenschaft konstruktiv zu begreifen. Es war und ist eine Ehre für mich, seine Schülerin zu sein.

Mein Dank gilt Prof. Reinhard Meister für die Freude an der Mathematik und Statistik, die er mir während und nach dem Studium stets vermittelt hat und für die statistische Beratung und Diskussion zu jeder Zeit.

Vielen Dank an Dr. Hans Knoblauch für die enge und sehr produktive wissenschaftliche Diskussion und Zusammenarbeit im Projektteam, für seine Energie, Arbeiten voranzutreiben und zu einem erfolgreichen Abschluss zu bringen. Seine kritische Durchsicht aus medizinischer Sicht hat mir sehr geholfen.

Vielen Dank an meine Kolleginnen Tatjana Luganskaja, für unzählige Rechnungen, die Datenaufbereitung und für ihre konstante mentale Unterstützung, und Katrin Hafez für wissenschaftliche Diskussionen und die kritische Durchsicht der Arbeit aus biochemischer Sicht.

Ganz großen Dank an meine Familie, an meine Kinder, für die Motivation und für die Sinnhaftigkeit dieses Vorhabens.

Berlin, den 3. Mai 2006

Anja Bauerfeind

Inhaltsverzeichnis

| | |
|------------------------------------------------------------------------|-----------|
| Vorbemerkung | v |
| Abkürzungen | ix |
| Symbole | xi |
| 1 Einleitung | 1 |
| 1.1 Stoffwechsel der Lipoproteine im menschlichen Blutplasma | 4 |
| 1.2 Kandidatengene, Funktion und Defekte | 6 |
| 1.3 Genetische Assoziationsstudien | 10 |
| Assoziation und Kopplung | 10 |
| Allelisches Spektrum komplexer Erkrankungen | 12 |
| Linkage disequilibrium | 15 |
| 1.4 Fragestellungen | 18 |
| 2 Methoden | 19 |
| 2.1 Studienpopulation | 19 |
| 2.2 Definition und Bestimmung der Phänotypen | 20 |
| 2.3 Definition und Bestimmung der SNP-Genotypen | 21 |
| 2.4 Analysestrategie | 21 |
| 2.5 Globale Analyse des Kopplungsungleichgewichts | 23 |
| 2.6 Nachweis phänotypischer Assoziation in Familien | 25 |
| 2.7 Vergleich genetischer Faktoren in zwei Populationen | 31 |
| 3 Ergebnisse | 35 |
| 3.1 Phänotypische Charakterisierung | 35 |
| 3.2 Genomische Karten und Kopplungsungleichgewicht | 36 |
| 3.3 Einfluss multipler SNPs und SNP-Haplotypen | 41 |
| 3.4 LD und Assoziation in zwei europäischen Populationen | 49 |
| 3.5 Populationsgenetischer Vergleich mit HapMap-Daten | 54 |

| | |
|----------------------------------------|------------|
| 4 Diskussion | 57 |
| 5 Zusammenfassung | 71 |
| Glossar | 73 |
| Literaturverzeichnis | 77 |
| A | 89 |
| A.1 Genomische Karten und LD | 89 |
| A.2 Haplotyp-Frequenzen | 98 |
| Curriculum Vitae | 107 |
| Publikationen | 109 |
| Erklärung | 111 |

Abkürzungen

Beschreibung der Abkürzungen in der Reihenfolge ihres Auftretens

| | |
|--------------|------------------------------------------------------------|
| FVK | Franz-Volhard-Klinik |
| MDC | Max-Delbrück-Centrum |
| LDL | Low density lipoprotein |
| HDL | High density lipoprotein |
| FH | Familiäre Hypercholesterinämie |
| SNP | Single nucleotide polymorphism |
| TC | Gesamtcholesterin |
| TG | Triglyzeride |
| LH-Ratio | Verhältnis: LDL/HDL |
| IDL | Intermediate density lipoprotein |
| VLDL | Very low density lipoprotein |
| CM | Chylomikronen |
| CE | Cholesterinester |
| <i>APOB</i> | Apolipoprotein B (Chromosom 2) |
| <i>APOE</i> | Apolipoprotein E (Chromosom 19) |
| <i>ABCA1</i> | ATP-binding cassette, sub-family A, member 1 (Chromosom 9) |
| <i>CETP</i> | Cholesteryl Ester Transfer Protein (Chromosom 16) |
| <i>LCAT</i> | Lecithin-Cholesteryl-Acyl-Transferase (Chromosom 16) |
| <i>LIPC</i> | Hepatische Triglyzeridlipase (Chromosom 15) |
| <i>LPL</i> | Hepatische Lipoproteinlipase (Chromosom 8) |
| <i>LDLR</i> | Low Density Lipoprotein Rezeptor (Chromosom 19) |
| <i>APOC2</i> | Apolipoprotein C-II (Chromosom 19) |
| <i>APOA5</i> | Apolipoprotein A-V (Chromosom 11) |
| <i>APOA4</i> | Apolipoprotein A-IV (Chromosom 11) |
| <i>APOC3</i> | Apolipoprotein C-III (Chromosom 11) |
| <i>APOA1</i> | Apolipoprotein A-I (Chromosom 11) |

| | |
|-------------|--------------------------------------------------|
| <i>PLTP</i> | Phospholipid Transferprotein (Chromosom 20) |
| <i>SRB1</i> | Scavenger Rezeptor, class B Typ I (Chromosom 12) |
| LD | Linkage disequilibrium |
| LE | Linkage equilibrium |
| CD/CV | Common Disease/Common Variants - Hypothese |
| BMI | Body Mass Index |
| AIC | Akaikes Informationskriterium |
| OR | Odds Ratio |
| KI | Konfidenzintervall |
| IBD | Identical By Descent |
| UTR | Untranslated region |

Symbole

Beschreibung häufig verwendeter Symbole in der Reihenfolge ihres Auftretens

| | |
|-------------------------------|-------------------------------------------------------------------------|
| N_e | Effektive Populationsgröße |
| p | Allelfrequenz |
| μ_t | Gesamtmutationsrate |
| μ_S | Neutrale Mutationsrate |
| μ_N | Reparaturrate |
| N | Neutrale Allele |
| S | Krankheitsverursachende Allele |
| ϕ | Allelische Identität |
| θ | Rekombinationswahrscheinlichkeit |
| c | Rekombinationsrate |
| r | Korrelationskoeffizient |
| $ r $ | Absoluter Wert des Korrelationskoeffizienten |
| Det | Determinante |
| β | Regressionsparameter |
| P | Quantitativer Phänotyp |
| G | Ggenotypische Effekte |
| E | Umweltbezogene Abweichung |
| μ | Erwartungswert |
| $\sigma^2(v)$ | Parameter (Schätzer) der phänotypischen Gesamtvarianz $Var(P)$ |
| $\sigma_{poly}^2(v_{poly})$ | Parameter (Schätzer) der genotypischen Varianz $Var(G)$ |
| $\sigma_{env}^2(v_{env})$ | Parameter (Schätzer) der Umweltvarianz $Var(E)$ |
| $\sigma_{locus}^2(v_{locus})$ | Parameter (Schätzer) der gemessenen genetischen Varianz eines Gen-Lokus |
| I | Einheitsmatrix |
| ϕ_{ij} | Kinship-Koeffizient zwischen zwei Individuen i und j |
| Ω | Varianz-Kovarianzmatrix |
| $\ln L$ | Log-Likelihood-Funktion |

D_{xy} Somers'D

R^2 Adjustiertes Bestimmtheitsmaß

Abbildungsverzeichnis

| | | |
|-----|---------------------------------------------------------------|----|
| 1.1 | Kinetisches Modell des Lipoproteinstoffwechsels | 5 |
| 1.2 | Beziehung zwischen Phänotyp und gemessener Variante | 11 |
| 3.1 | Histogramm: LDL- und HDL-Cholesterin | 40 |
| 3.2 | Genomische Karte: <i>APOB</i> | 40 |
| 3.3 | Genomische Karte: <i>LIPC</i> | 41 |
| 3.4 | Kopplungsungleichgewicht (LD) im <i>APOB</i> -Gen | 42 |
| 3.5 | Kopplungsungleichgewicht (LD) im <i>LIPC</i> -Gen | 43 |
| 3.6 | Lokus-Anteile an genetisch bedingter LDL-Varianz | 48 |
| 3.7 | Lokus-Anteile an genetisch bedingter HDL-Varianz | 48 |
| 3.8 | Allelfrequenzen: Berlin versus HapMap | 55 |
| 3.9 | Linkage Disequilibrium: Berlin versus HapMap | 55 |
| A.1 | Genomische Karte und LD im <i>ABCA1</i> -Gen | 90 |
| A.2 | Genomische Karte und LD im <i>APOE</i> -Gen | 91 |
| A.3 | Genomische Karte und LD im <i>CETP</i> -Gen | 92 |
| A.4 | Genomische Karte und LD im <i>LCAT</i> -Gen | 93 |
| A.5 | Genomische Karte und LD im <i>LPL</i> -Gen | 94 |
| A.6 | Genomische Karte und LD im <i>LDLR</i> -Gen | 95 |
| A.7 | Genomische Karte und LD im <i>APOC2</i> -Gen | 96 |
| A.8 | Genomische Karte und LD im Gencluster | 97 |

Tabellenverzeichnis

| | | |
|------|----------------------------------------------------------------------------|-----|
| 3.1 | Verteilung der Lipid-Konzentrationen | 36 |
| 3.2 | Kandidatengene und Genetische Marker | 39 |
| 3.3 | Varianzkomponentenzerlegung im Gesamtmodell | 44 |
| 3.4 | Block-Bootstrap-KI der Varianzkomponenten | 45 |
| 3.5 | Signifikanz des partiellen Lokus-Beitrags | 46 |
| 3.6 | Geschlechtsspezifische Signifikanz des partiellen Lokus-Beitrags | 47 |
| 3.7 | Genomische Distanz untersuchter SNPs: Berlin/Genf | 49 |
| 3.8 | Allelfrequenzen: Berlin/Genf | 50 |
| 3.9 | Linkage Disequilibrium: Berlin/Genf | 51 |
| 3.10 | SNP Assoziation: Berlin/Genf | 53 |
| 3.11 | Gen-Lokus Effekte: Berlin/Genf | 54 |
| A.1 | Haplotyp-Frequenzen | 105 |

Curriculum Vitae

Mein Lebenslauf wird aus Datenschutzgründen in der elektronischen Version meiner Arbeit nicht mit veröffentlicht.

Publikationen

Relevante eigene Publikationen

- [1] Bauerfeind A, Knoblauch H, Costanza MC, Luganskaja T, Toliat MR, Nürnberg P, Luft FC, Reich JG & Morabia A. Concordant association of lipid gene variation with a combined HDL/LDL-cholesterol phenotype in two European populations. *Hum Hered* 2006; in press.
- [2] Nürnberg P, Reich JG, Luft FC, Knoblauch H, Bauerfeind A, Luganskaja T, Rohde K. System von SNP-Haplotypen des menschlichen Lipoproteinstoffwechsels zur Diagnose und Prognose des Lipidstatus des Menschen. Europäisches Patentamt 2004.
- [3] Knoblauch H, Bauerfeind A, Toliat MR, Becker C, Luganskaja T, Günther UP, Rohde K, Schuster H, Junghans C, Luft FC, Nürnberg P & Reich JG. Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet* 2004;13:993–1004.
- [4] Bauerfeind A, Knoblauch H, Schuster H, Luft FC & Reich JG. Single nucleotide polymorphism haplotypes in the cholesteryl-ester transfer protein (CETP) gene and lipid phenotypes. *Hum Hered* 2002;54:166–173.
- [5] Knoblauch H, Bauerfeind A, Krahenbuhl C, Daury A, Rohde K, Bejanin S, Essioux L, Schuster H, Luft FC & Reich JG. Common haplotypes in five genes influence genetic variance of LDL and HDL cholesterol in the general population. *Hum Mol Genet* 2002;11:1477–1485.

Vorträge

- 11/2004 *Modellierung der Genotyp-Phänotyp-Assoziation im Lipoproteinstoffwechsel auf der Basis von Großfamilien*
Biometrisches Kolloquium, Humboldt-Universität zu Berlin

03/2004 *Haplotypes and SNPs in 13 lipid-relevant genes: Contribution to genetic variance ob lipid traits*

Workshop „Analysis of Complex Genetic Traits“ Universität Rostock, Medizinische Fakultät

11/2002 *Common haplotypes in five genes influence genetic variance ob LDL and HDL cholesterol in the general population*

NGFN/ DHGP Symposium 2002, Berlin

Poster

04/2004 *Haplotypes and SNPs in 13 lipid-relevant genes*

Human Genome Meeting 2004, Berlin

03/2004 *Contribution to genetic variance ob lipid traits*

European Mathematical Genetics Meeting (EMGM) 2004, Kiel

10/2002 *Genetic influence on LDL- and HDL-Cholesterol*

Jahrestagung der Deutschen Gesellschaft für Humangenetik 2002, Leipzig

Weitere eigene Publikationen

- [1] Follmann M, Heinemann LAJ, Bauerfeind A & Garbe E. Treatment with Potentially Hepatotoxic Drugs and the Risk ob Hepatocellular Carcinoma: Results ob a European Case - Control Study. *Pharmacoepidemiology and drug safety* 2000;9:417–422.
- [2] Strohmeyer D, Rössing C, Strauss F, Bauerfeind A, Kaufman O & Loening S. Tumor Angiogenesis Is Associated With Progression After Radical Prostatectomy in p T2/p T3 Prostate Cancer. *The Prostate* 2000;42(1):26–33.
- [3] Strohmeyer D, Rössing C, Bauerfeind A, Kaufman O, Schlechte H, Bartsch G & Loening S. Vascular endothelial Growth Factor and Its Correlation With Angiogenesis and p53 Expression in Prostate Cancer. *The Prostate* 2000;45(3):216–224.
- [4] Mantay K, Bauerfeind A, Suissa S & Garbe E. Risk ob serious hypoglycaemia and use ob antihypertensive drugs in a cohort ob diabetic patients treated with insulin. *Pharmacoepidemiol Drug Saf* 1999;8(Suppl.2):257.
- [5] Mantay K, Bauerfeind A, Suissa S & Garbe E. Antihypertensive drug treatment and the risk ob severe hypoglycaemia in patients treated with oral antidiabetics. *Pharmacoepidemiol Drug Saf* 1999;8(Suppl.2):258.

Erklärung

Ich, Anja Bauerfeind, erkläre, dass ich die vorgelegte Dissertationsschrift mit dem Thema: *Einfluss genetischer Varianten auf den Cholesterinstoffwechsel in einer mitteleuropäischen Bevölkerungsstichprobe* selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, ohne die (unzulässige) Hilfe Dritter verfasst und auch in Teilen keine Kopien anderer Arbeiten dargestellt habe.

Berlin, den 3. Mai 2006

Anja Bauerfeind