



# **BMBF-Projekt**

## **„Qualitätsmanagement für Hochdurchsatz-Genotypisierung“**

*Statusbericht TP 2 (Stats)*

*"Informationsveranstaltung "Qualitätsmanagement für Hochdurchsatz-Genotypisierung", 21. Juni 2010, Berlin*

*Bertram Müller-Myhsok, MPI Psychiatrie, München*

---

GEFÖRDERT VOM



## Grundlegende Idee

---

- **Qualitätskontrollparameter werden oftmals nach „Bauchgefühl“ beurteilt und gesetzt**
- **Daher: Untersuchung eines bestimmten „Endpunkts“ (zB genomic inflation factor lambda) unter verschiedenen Parameterwerten für typische Qualitätskontrollparameter (MAF, HWE [p], CR[SNP], CR[ID])**



## Datensätze

---

- **Illumina (~810 cases, ~840 controls, Illumina 550, Illumina 610)**
  - Echter Phänotyp
  - Permutierter Phänotyp
- **Affymetrix (400 Fälle, 1644 controls, Affy 500k, RLS, Winkelmann et al. Nat Genet 2007)**
  - Echter Phänotyp
  - Permutierter Phänotyp



# Vorgehen

```
#!/bin/bash
# script for TMF

MAF='0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.11 0.12 0.13 0.14 0.15'
HWE='0.05 0.01 0.001 0.0001 0.00001'
CR='0.50 0.52 0.54 0.56 0.58 0.60 0.62 0.64 0.66 0.68 0.70 0.72 0.74 0.76 0.78 0.80 0.82 0.84 0.86 0.88 0.90
0.92 0.94 0.96 0.98 1.00'

for mind in `echo $CR`; do
    for geno in `echo $CR`; do
        for hardy in `echo $HWE`; do
            for minor in `echo $MAF`; do
                #echo $geno;
                #echo $mind;
                #echo $hardy;
                #echo $minor;
                plink --pheno-name CC --bfile rls --pheno
                rls.phen --adjust --out rls_${minor}_${hardy}_${mind}_${geno} --assoc --geno $geno --mind $mind --maf
                $minor --hwe $hardy
                rm *assoc;
                rm *adjusted;
            done;
        done;
    done;
done;
```



## Procedure (2)

---

- **60840 Analysen pro Experiment**
- **Rechenzeit PLINK:**
  - 3 mins pro Analyse
  - 126 days pro Experiment
- **GenABEL:**
  - Rechenzeit um Faktor 30 reduziert.



## Aus beiden ...

---

- **Illumina (~810 Fälle, ~840 Kontrollen, Illumina 550, Illumina 610)**
  - Echter Phänotyp
  - Permutierter Phänotyp
- **Affymetrix (400 cases, 1644 Kontrollen, Affy 500k, RLS, Winkelmann et al. Nat Genet 2007)**
  - True phenotype
  - Permuted phenotype
- **Generell: Permutierte Phänotypen ohne große Hinweise auf geeignete Parameterwerte (kein relevanter gerichteter Fehler)**
- **Anders: echter Phänotyp**

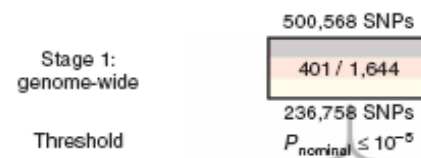
# Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions

Juliane Winkelmann<sup>1-3</sup>, Barbara Schormair<sup>1,3</sup>, Peter Lichtner<sup>1,3</sup>, Stephan Ripke<sup>2</sup>, Lan Xiong<sup>4</sup>, Shapour Jalilzadeh<sup>1,3</sup>, Stephany Fulda<sup>2</sup>, Benno Pütz<sup>2</sup>, Gertrud Eckstein<sup>1,3</sup>, Stephanie Hauk<sup>1,3</sup>, Claudia Trenkwalder<sup>5</sup>, Alexander Zimprich<sup>6</sup>, Karin Stiasny-Kolster<sup>7</sup>, Wolfgang Oertel<sup>7</sup>, Cornelius G Bachmann<sup>8</sup>, Walter Paulus<sup>8</sup>, Ines Peglau<sup>9</sup>, Ilonka Eisensehr<sup>10</sup>, Jacques Montplaisir<sup>11,12</sup>, Gustavo Turecki<sup>13</sup>, Guy Rouleau<sup>4</sup>, Christian Gieger<sup>14</sup>, Thomas Illig<sup>14</sup>, H-Erich Wichmann<sup>14,15</sup>, Florian Holsboer<sup>2</sup>, Bertram Müller-Myhsok<sup>2,16</sup> & Thomas Meitinger<sup>1,3,16</sup>

**Restless legs syndrome (RLS) is a frequent neurological disorder characterized by an imperative urge to move the legs during night, unpleasant sensation in the lower limbs, disturbed sleep and increased cardiovascular morbidity. In a genome-wide association study we found highly significant associations between RLS and intronic variants in the homeobox gene *MEIS1*, the *BTBD9* gene encoding a BTB(POZ) domain as well as variants in a third locus containing the genes encoding mitogen-activated protein kinase *MAP2K5* and the transcription factor *LBXCOR1* on chromosomes 2p, 6p and 15q, respectively. Two independent replications confirmed these association signals. Each genetic variant was associated with a more than 50% increase in risk for RLS, with the combined allelic variants conferring more than half of the risk. *MEIS1* has been implicated in limb development, raising the possibility**

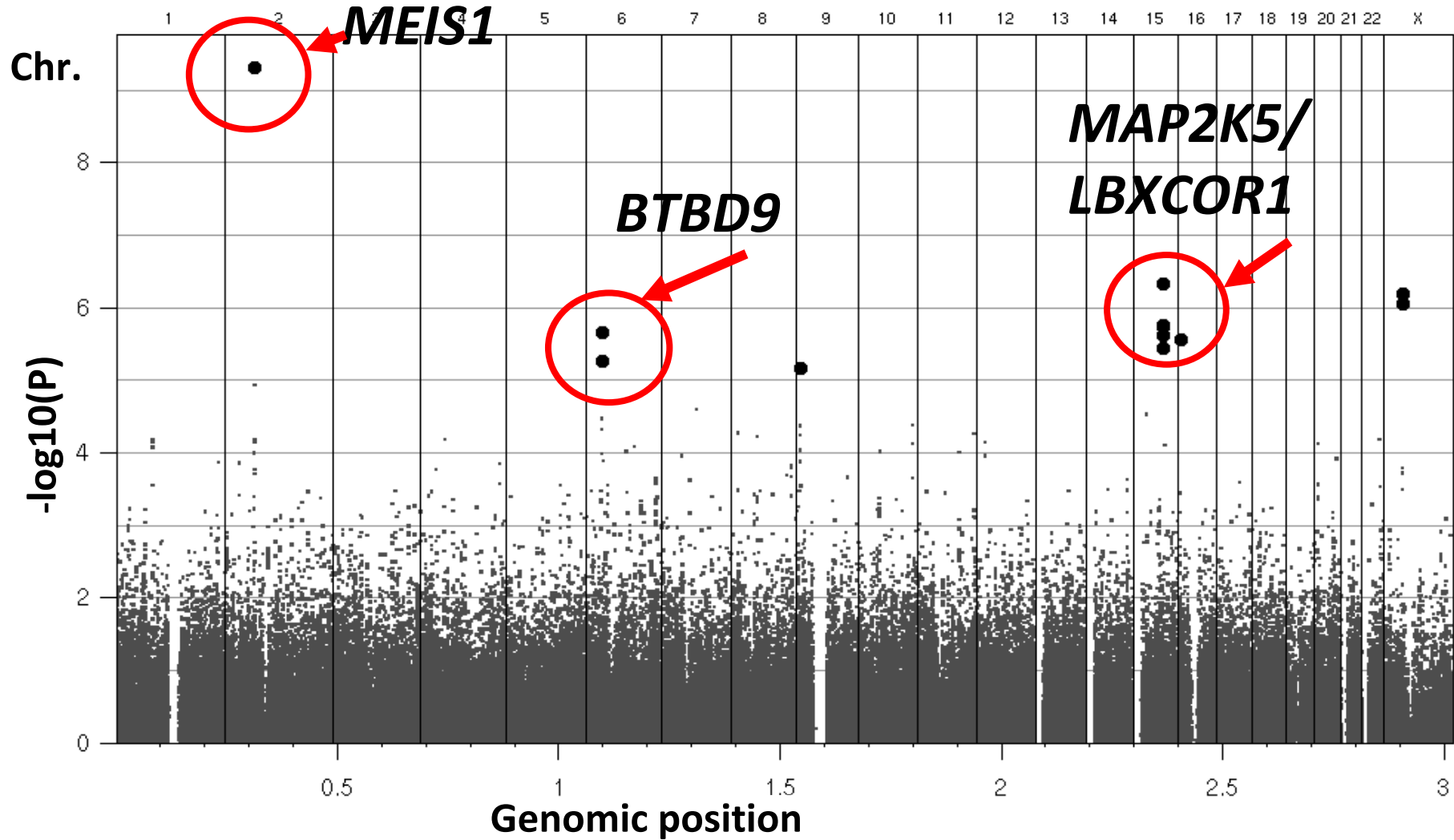
– that RLS has components of a developmental disorder.

prevalence of up to 10% in the elderly in North America and Europe<sup>2</sup>. Dopaminergic agents originally developed for Parkinson's disease have been used to treat RLS, with an unknown mode of action<sup>2</sup>. Neurophysiological, pharmacological and neuroimaging studies suggest that the characteristic symptoms originate in the central nervous system, yet the underlying neurobiology remains obscure<sup>4</sup>. A family history of RLS is present in more than 50% of affected individuals, and similar figures have been reported for heritability in twin studies<sup>5,6</sup>. Linkage analysis uncovered five loci based on recessive (RLS1) or dominant inheritance (RLS2–RLS5), but so far





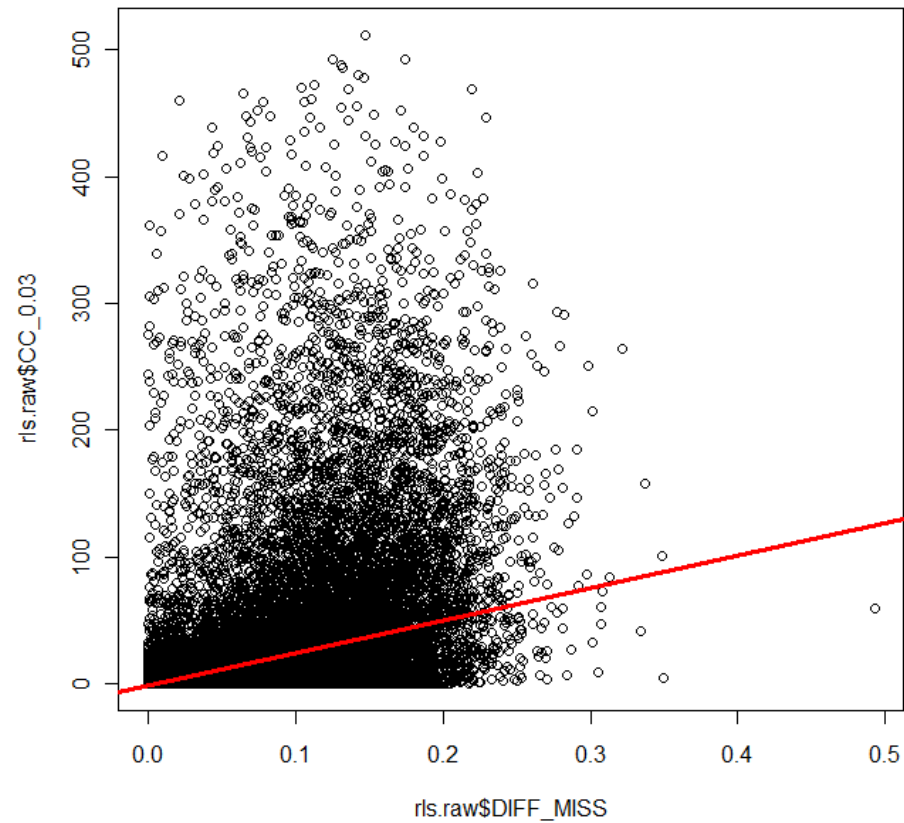
# Genomewide Association Study







## Well known things ...



GEFÖRDERT VOM



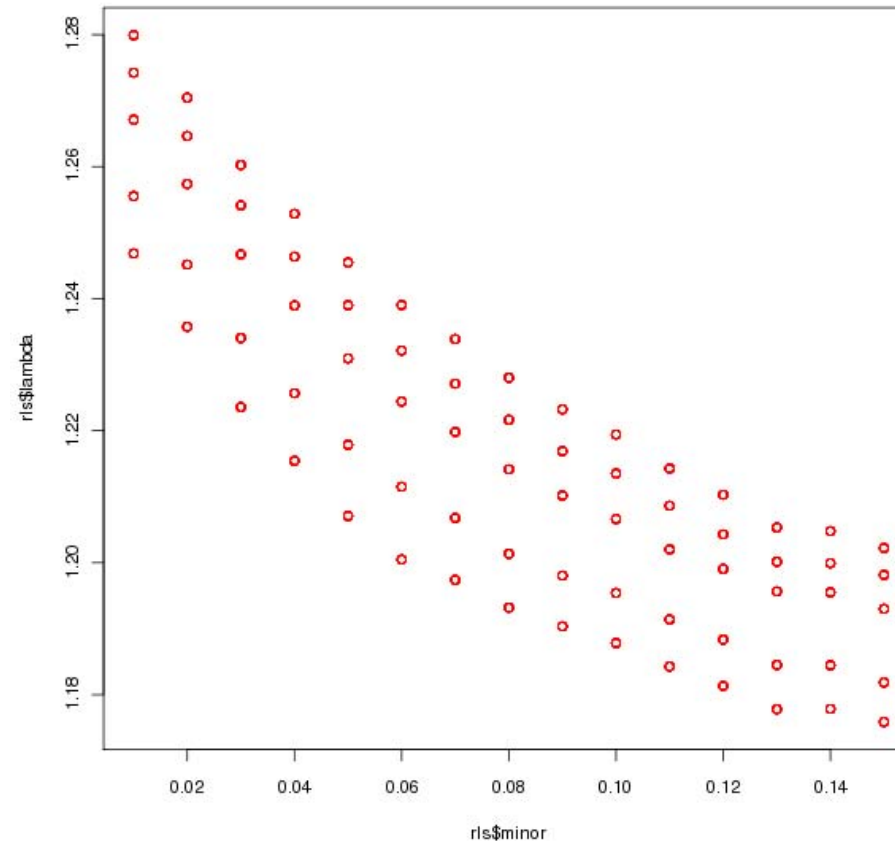
Bundesministerium  
für Bildung  
und Forschung

Qualitätsmanagement für Hochdurchsatz-Genotypisierung  
Teilprojekt 2 - Plausibilitätskriterien

21.06.2010  
Folie 9



## Deutliche Abhängigkeiten ...



GEFÖRDERT VOM



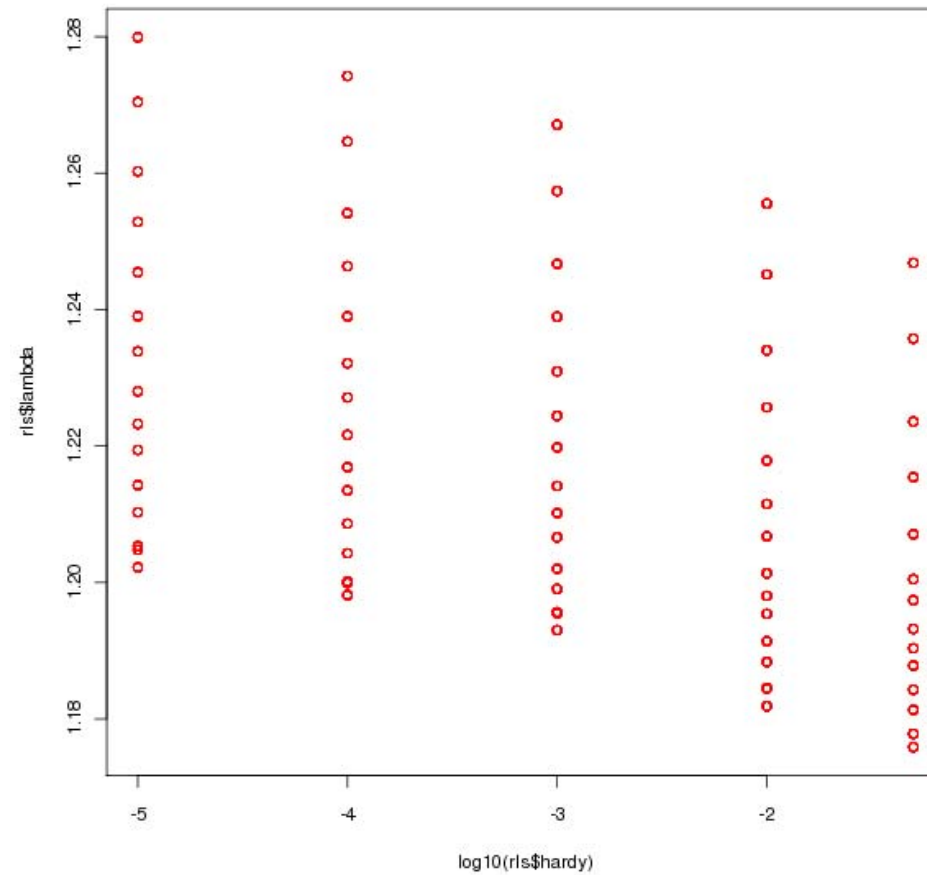
Bundesministerium  
für Bildung  
und Forschung

Qualitätsmanagement für Hochdurchsatz-Genotypisierung  
Teilprojekt 2 - Plausibilitätskriterien

21.06.2010  
Folie 10



## Deutliche Abhängigkeiten ...



GEFÖRDERT VOM



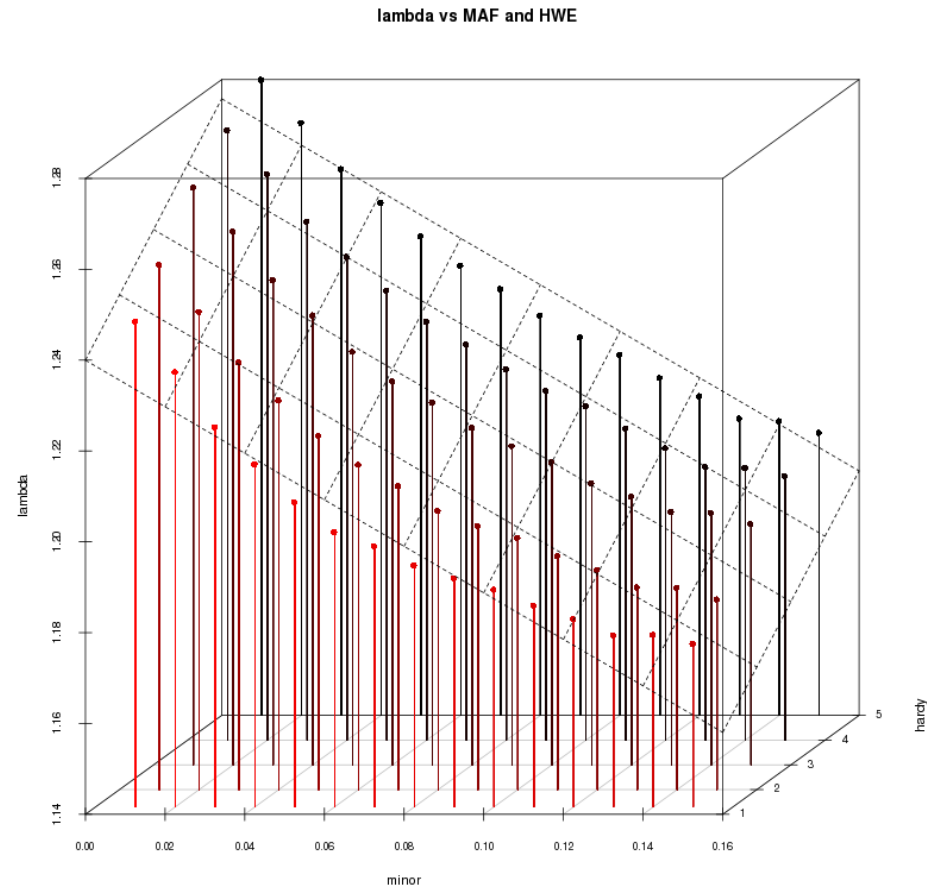
Bundesministerium  
für Bildung  
und Forschung

Qualitätsmanagement für Hochdurchsatz-Genotypisierung  
Teilprojekt 2 - Plausibilitätskriterien

21.06.2010  
Folie 11



# Deutliche Abhängigkeiten ...



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

Qualitätsmanagement für Hochdurchsatz-Genotypisierung  
Teilprojekt 2 - Plausibilitätskriterien

21.06.2010  
Folie 12



## Effekte sind nicht nur additiv....

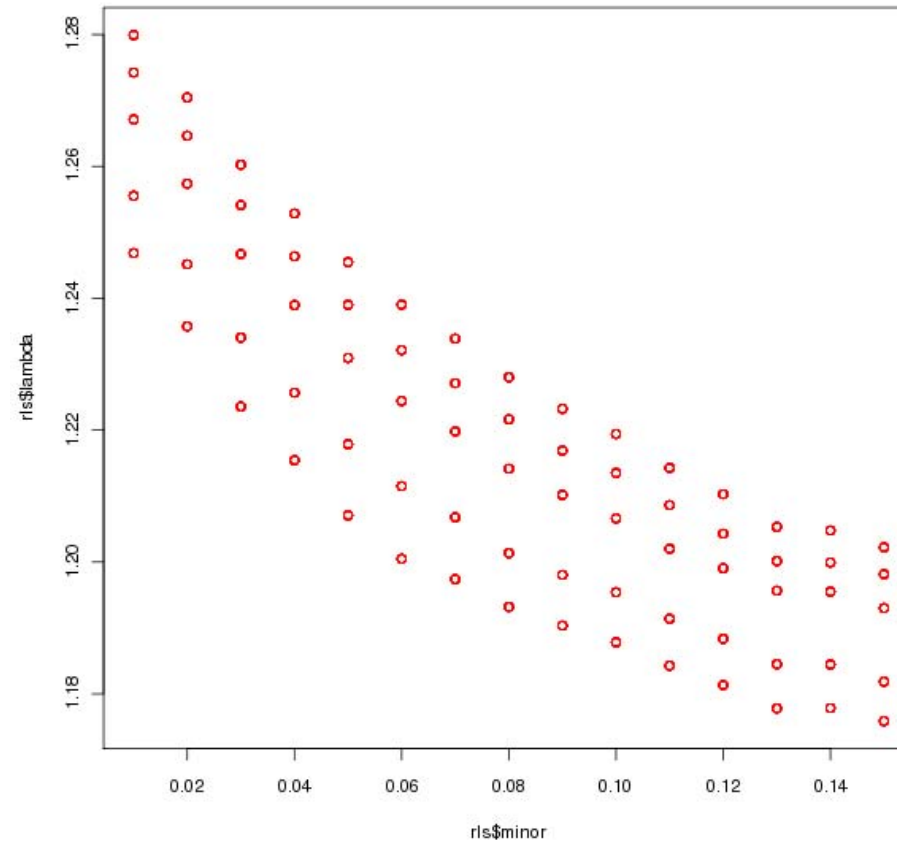
```
anova(step(lm(rls.raw$CC_0.03~(rls.raw$ALL_MISS+rls.raw$DIFF_MISS+rls.raw$MAF+rls.raw$HWE)^2))  
,test="F")
```

Response: rls.raw\$CC\_0.03

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
rls.raw\$ALL_MISS	1	15720150	15720150	104108.442	< 2.2e-16	***
rls.raw\$DIFF_MISS	1	8454214	8454214	55988.971	< 2.2e-16	***
rls.raw\$MAF	1	97528	97528	645.888	< 2.2e-16	***
rls.raw\$HWE	1	4007	4007	26.535	2.588e-07	***
rls.raw\$ALL_MISS:rls.raw\$DIFF_MISS	1	12082754	12082754	80019.380	< 2.2e-16	***
rls.raw\$ALL_MISS:rls.raw\$MAF	1	2037105	2037105	13490.957	< 2.2e-16	***
rls.raw\$ALL_MISS:rls.raw\$HWE	1	778955	778955	5158.715	< 2.2e-16	***
rls.raw\$DIFF_MISS:rls.raw\$MAF	1	2501504	2501504	16566.488	< 2.2e-16	***
rls.raw\$DIFF_MISS:rls.raw\$HWE	1	71036	71036	470.441	< 2.2e-16	***



# Was kann man lernen?



GEFÖRDERT VOM



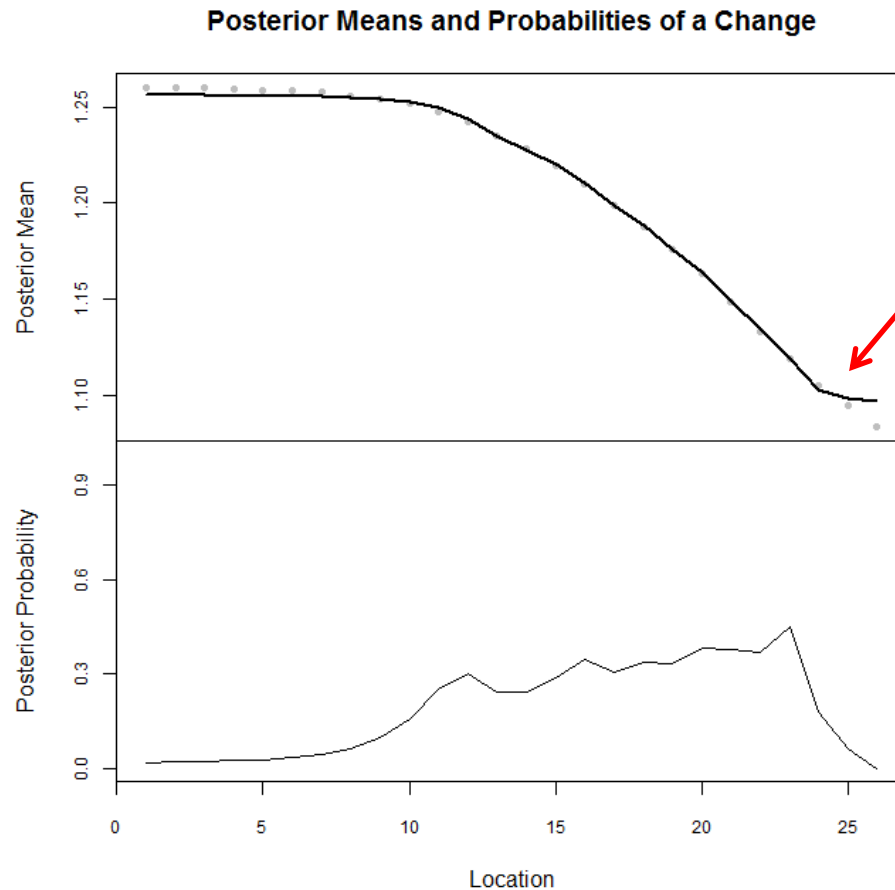
Bundesministerium  
für Bildung  
und Forschung

Qualitätsmanagement für Hochdurchsatz-Genotypisierung  
Teilprojekt 2 - Plausibilitätskriterien

21.06.2010  
Folie 14



# Aus den Daten selbst?



SNP CR  
0.97

Change point analysis  
(R package bcp V 2.1.3),  
Parameter varying here  
Call rate per SNP, other  
parameters:  
Ind CR 0.96, HWE  $1e-5$ ,  
MAF 0.02

GEFÖRDERT VOM



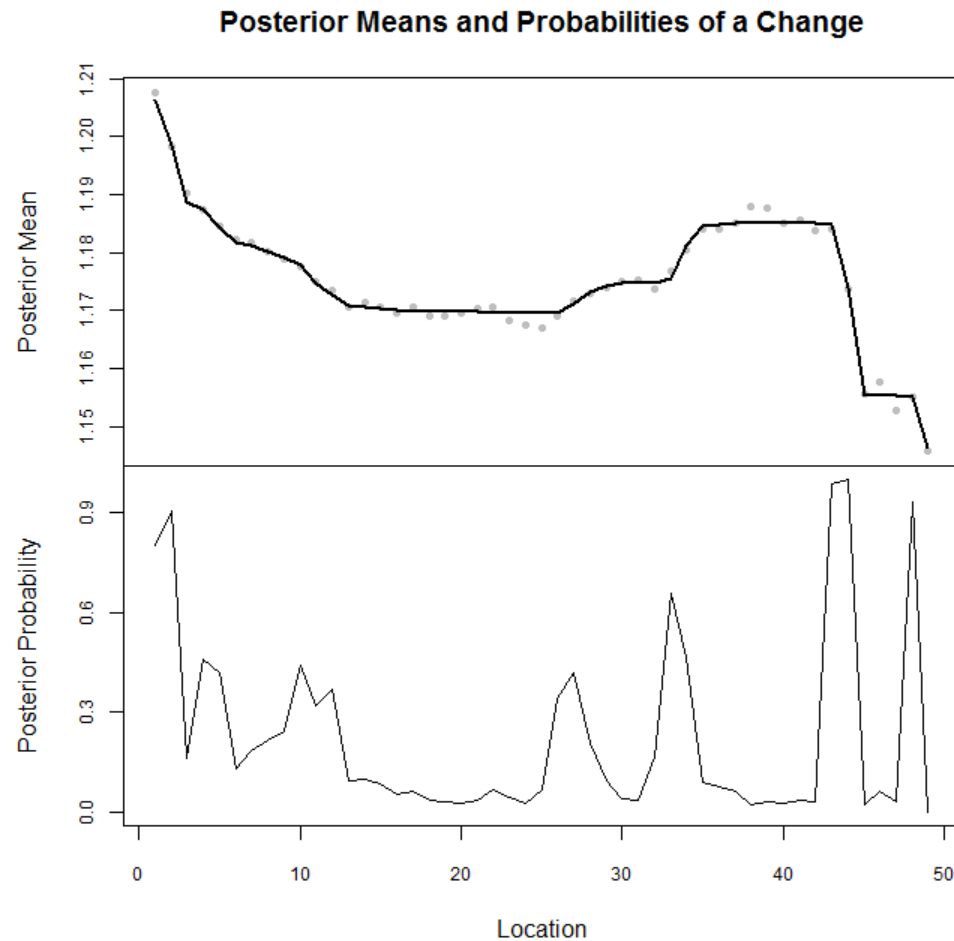
Bundesministerium  
für Bildung  
und Forschung

Qualitätsmanagement für Hochdurchsatz-Genotypisierung  
Teilprojekt 2 - Plausibilitätskriterien

21.06.2010  
Folie 15



# Die Daten selbst gehen gelegentlich eine deutliche Antwort



Change point analysis  
(R package bcp V 2.1.3),  
Parameter varying here  
MAF (from 0.01 to 0.49,  
other parameters:  
Ind CR 0.95, HWE  $1e-5$ ,  
SNP CR 0.90

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

Qualitätsmanagement für Hochdurchsatz-Genotypisierung  
Teilprojekt 2 - Plausibilitätskriterien

21.06.2010  
Folie 16





# Zusammenfassung

---

- **Permutierte Phänotypen: keine systematischen Fehler**
- **Echte Phänotypen – Möglichkeit systematischer Fehler:**
  - Alle Parameter können eine Rolle spielen
  - Evtl Information aus den Daten selbst (change point Analysis)
  - **Rechenintensiv**
  - GPGPU?