Aus dem Max Planck Institut für molekulare Genetik

DISSERTATION

Development and application of CGHPRO, a novel software package for retrieving, handling and analysing array CGH data

Zur Erlangung des akademischen Grades
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

Vorgelegt von

Wei Chen

aus Fujian, V. R. China

Juli, 2006

1st Reviewer: **Prof. Dr. H.-Hilger Ropers**

2nd Reviewer: **Prof. Dr. Gerd Multhaup**

Date of defence: Oct. 27, 2006

Acknowledgment

First, I would like to thank Prof. Dr. Hans-Hilger Ropers for guiding me into the fascinating research field of Human Genetics, and for his supervision throughout my project. I benefit from the discussion with him very much, which is of vital importance for my future career development. I am indebted to Professor Dr. Gerd Multhaup for being my adviser in Free University Berlin.

I would like to thank Dr. Reinhard Ullmann for stimulating discussions and valuable suggestions, which helped in understanding the technical details in molecular cytogenetic experiments, and for giving critical comments on my thesis.

Dr. Andreas W. Kuss deserves my thanks for his expert advice, and for thoroughly reading my thesis.

.

Thanks are also due to members of Dr. Ullmann's group, especially Dr. Fikret Erdogan, whose experimental work, which produced the basis for my work.

I would like to acknowledge my colleagues, Dr. Lars Riff Jensen, Dr. Andreas Tzschach, Dr. Steffen Lezner, Melanie Wendehack, Masoud Gashasbi, Mohammad Mahdi Motazacker, Lia Abbasi-Moheb, Bettina Lipkowitz, Marianne Schlicht and Marion Amende-Acar, for creating a pleasant working atmosphere throughout my project.

Finally, I want to express my deep appreciation to my families. My wife, Yuhui, her continuous love, care and sweetness has made my life full of joy and interests. My parents, my parents-in-law, and my sister, only your continuous support over all these years enabled me to get this far.

Summary

Genome rearrangements contribute significantly to the etiology of genetic disorders but also to human genetic diversity and disease susceptibility. For the detection of submicroscopic deletions and duplications on a genome wide level, a BAC-Array based technique for comparative genomic hybridisation (Array CGH), using a high number of overlapping BACs covering the whole genome is now being applied. The resulting data output however is of a magnitude that requires powerful management tools for handling not only large data quantities but also for coping with data quality variation.

To facilitate the analysis and management of array CGH data, I have developed a comprehensive software package called 'CGHPRO'. Using the results from the image analysis software, CGHPRO allows hybridisation features to be checked with a variety of graphical representation options, thus enabling the selection of the most suitable normalisation method for individual experiments. A variety of options is then offered to characterize individual genomic profiles from the normalized data sets. All results are visualized in an interactive interface and stored in a database. The database allows the repetitive use of the stored results in comparative analyses, e.g. for investigating chromosomal aberration patterns in specific patient cohorts. In order to take the resolution of ArrayCGH applications beyond the BAC level CGHPRO allows the design of high-resolution specific sub-arrays.

The power of CGHPRO was demonstrated in the analysis of 22 mentally retarded patients with submegabase resolution whole genome tiling path BAC array CGH, which led to the identification of 20 deletions and two duplications. Additionally, as a proof of principle for CGHPRO assisted sub-array design, the breakpoints from a balanced translocation t(1;13) were successfully fine mapped.

When comparing the breakpoint regions for the 22 mentally retarded patients with those from a set of 41 balanced translocation carriers, in 6 of 22 unbalanced aberrations, breakpoint flanking duplications with a high degree of sequence

similarity were found, suggesting that unequal crossing over might be one factor in chromosome instability. In all 41 balanced translocations however, even though breakpoint flanking duplications were observed, sequence homology between them never occured. This second finding indicates the existence of additional chromosomal instability factors which depend on or coincide with segmental duplications.

Taken Together, the results presented here demonstrate the powerful enhancement of the Array-CGH technique by the development and application of a versatile data management and ananlysis tool. It can be concluded, that the implementation of the protocols introduced here will, also for studies in large patient cohorts, greatly facilitate the identification and investigation of disease-associated chromosomal aberrations.

.

Contents

1 Preface 1
2 Genome rearrangement 2
3 Molecular cytogenetics 4
3.1 Fluorescence in situ hybridisation (FISH) 4
3.2 Comparative Genomic Hybridisation (CGH) and array CGH 5
3.2.1 Experimental platforms for array CGH 8
3.2.2 Data Analysis of array CGH 12
4 Development of CGHPRO 15
4.1 Software development 15
4.2 Notation of array CGH data 15
4.2.1 Intensities 16
4.2.2 Ratios 16
4.2.3 Log-intensities and log-ratios 16
4.3 Overview of the data analysis process in CGHPRO 16
4.4 Database design 17
4.5 Data input 18
4.6 Graphical analysis of hybridisation characteristics 19
4.6.1 Scatter plot 19
4.6.2 MA-plot 19
4.6.3 Boxplot 21
4.6.4 Histogram 21
4.7 Data filter 22
4.8 Data normalization 22
4.8.1 Within-slide normalization 23
4.8.1.1 Global normalization 23
4.8.1.2 Intensity dependent normalization 23
4.8.1.3 Printing tip specific normalization 23
4.8.2 Dye-Swap normalization 24
4.9 Replicate spots handling 25
4.10 Characterization of genomic profiles 25
4.11 Data display 26
4.11.1 Genomic display 26
4.11.2 Chromosome display 29
4.12 Comparative analysis of different chips 29
4.13 Application of CGHPRO in sub-array design 31
4.14 Batch analysis 31
4.15 Availability of CGHPRO 31
5 Application of array CGH and CGHPRO 32
5.1 Impact of segmental duplication on the generation of genome rearrangements . 32
5.1.1 Molecular mechanisms underlying genome rearrangements 32
5.1.1.1 Segmental duplication-mediated nonallelic homologous recombination 33
5.1.1.2 Other genome architectural features 34

5.1.2 Copy number changes in 22 patients with mental retardation 35
5.1.3 Overlap of breakpoints in unbalanced aberrations with segmental duplication and CNPs 39
5.1.4 Overlap of breakpoints in balanced translocation with segmental duplication and CNPs 44
5.1.5 Segmental duplication and CNPs in balanced and unbalanced rearrangements 46
5.1.6 Segmental duplication could mediate non-allelic homologous recombination (NAHR) 48
5.1.7 NAHR involving segmental duplications cannot explain all unbalanced rearrangements 49
5.1.8 NAHR involving segmental duplications cannot explain balanced rearrangements 50
5.1.9 Clusters of segmental duplications may cause chromosomal instability 51
5.2 Mapping of balanced translocation breakpoints using array CGH and CGHPRO.. 52
5.2.1 Cytogenetic investigation of the translocation 53
5.2.2 Array CGH investigation of the translocation 54
5.2.3 Confirmation of the chromosome breakpoints by FISH analysis 55
5.2.4 PCR fragment subarray 55
5.2.5 Long-range PCR and sequencing 56
5.2.6 Advantage of array painting in mapping balanced translocation breakpoints 58
6 Summary 59
7 Zusammenfassung 61
8 Reference 63
9 Supplemental material 70
9.1 Experimental protocols 70
9.1.1 Array CGH 70
9.1.1.1 Array production 72
9.1.1.2 DNA labelling and hybridisation 74
9.1.1.3 Scanning 75
9.1.2 Mapping balanced translocation breakpoint by chromosome sorting and array painting 75
9.1.2.1 Cell culture and preparation 75
9.1.2.2 Flow karyotyping and sorting 76
9.1.2.3 DNA amplification using GenomePlex 76
9.1.2.4 Array painting using 36K BAC array 76
9.1.2.5 Fluorescence in situ hybridisation 76
9.1.2.6 PCR fragment subarray 77
9.1.2.7 Long-range PCR and sequencing 82
9.1.2.8 Agarose gel electrophoresis 83
9.2 Supplementary tables 84
***Curriculum Vitae* 95**