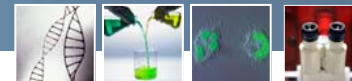


Copy Number Variation Analysis Using Quantitative TaqMan[®] Copy Number Assays

**Kelly Li, Ph.D.
Senior Staff Scientist
Genomic Assays R&D**

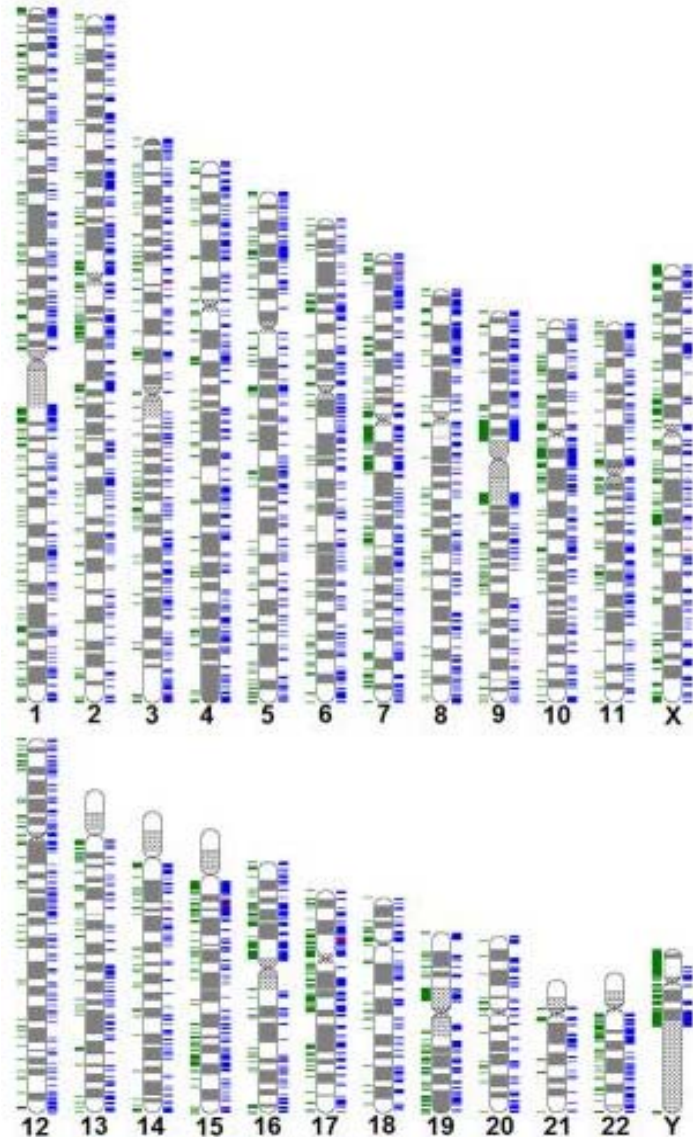
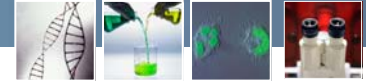
**Shanghai, China
Nov. 26, 2009**





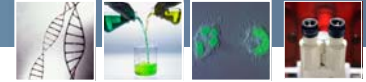
- Copy number variation (CNV)
- TaqMan® Copy Number Assays
- Application studies of TaqMan® Copy Number Assays
 - Validation of genome-wide CNV discovery
 - Large sample screening
 - Clarification of discrepancy from other platforms
 - Isoform-specific copy number detection: C4 family
 - Exon-specific deletion detection: OTC gene
- Summary

Copy Number Variation (CNV)



- Copy Number Variants (CNVs)
 - A segment of DNA that is 1kb or larger and is present at variable copy number in comparison with a reference genome
- High frequency and significance of CNV
 - Much more common than expected
 - 5-20% of human genome is CNVR
 - Normal polymorphism and disease-associated
 - Autoimmune diseases
e.g. SLE, Psoriasis, Crohn Disease
 - Neurological diseases
e.g. Autism, Schizophrenia, Bipolar
 - Genetic/Genomic disorders: e.g. microdeletion/microduplication syndromes
 - HIV infection
 - Cancer
- CNV is increasingly being recognized as
 - a substantial fraction of total genetic variability
 - an important part of genome structural variation

Current Technologies and Challenges for CNV Detection



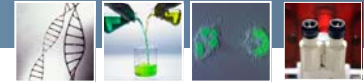
➤ Current technologies

- Array-based: whole genome CNV discovery
- FISH: clinical diagnostics
- NextGen sequencing
- Ligation-based: MLPA
- QPCR-based: e.g. SYBR and TaqMan

➤ Needs/challenges from the CNV field

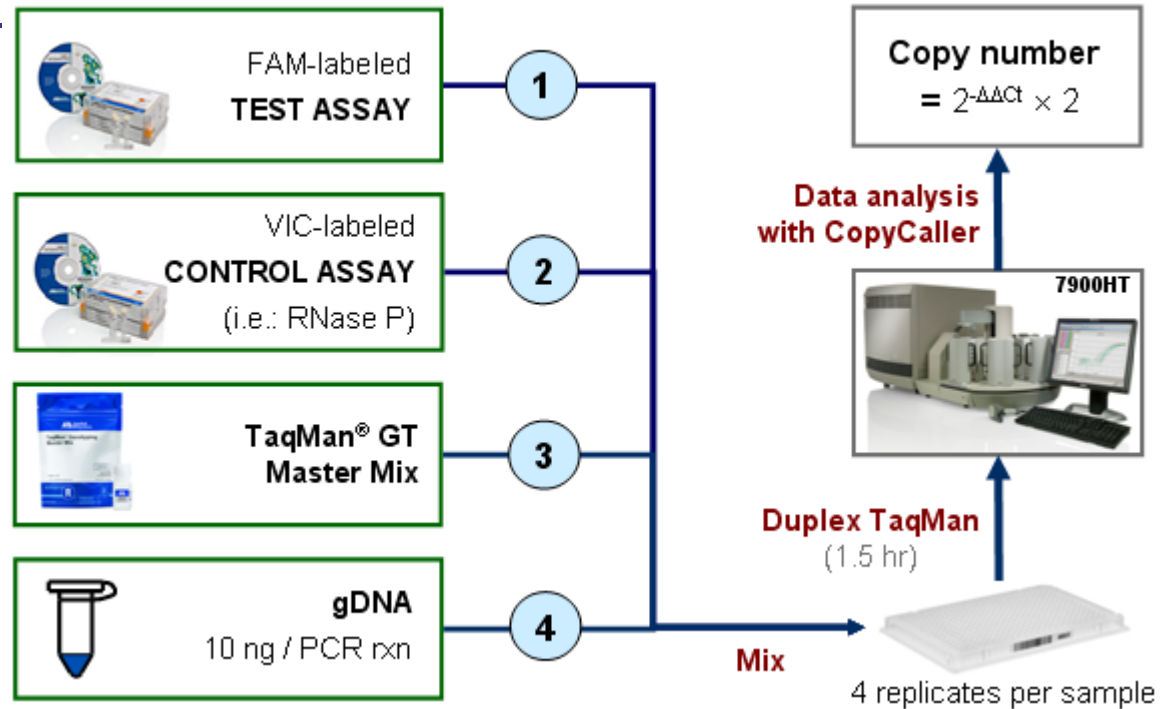
- reliable validation tools for genome-wide CNV discovery from different platforms
- more quantitative, accurate and specific
- simple workflow
- sample throughput for known targets

TaqMan Copy Number Assays

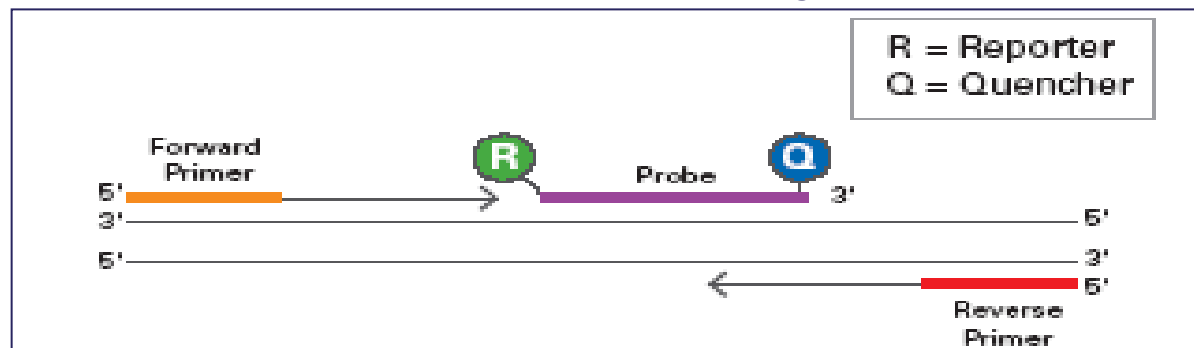


TaqMan Copy Number Assay

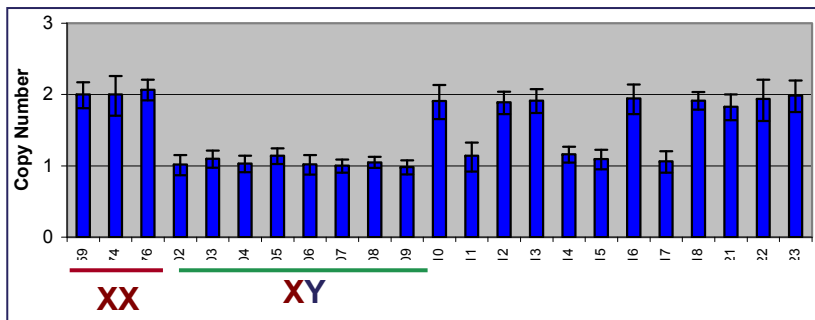
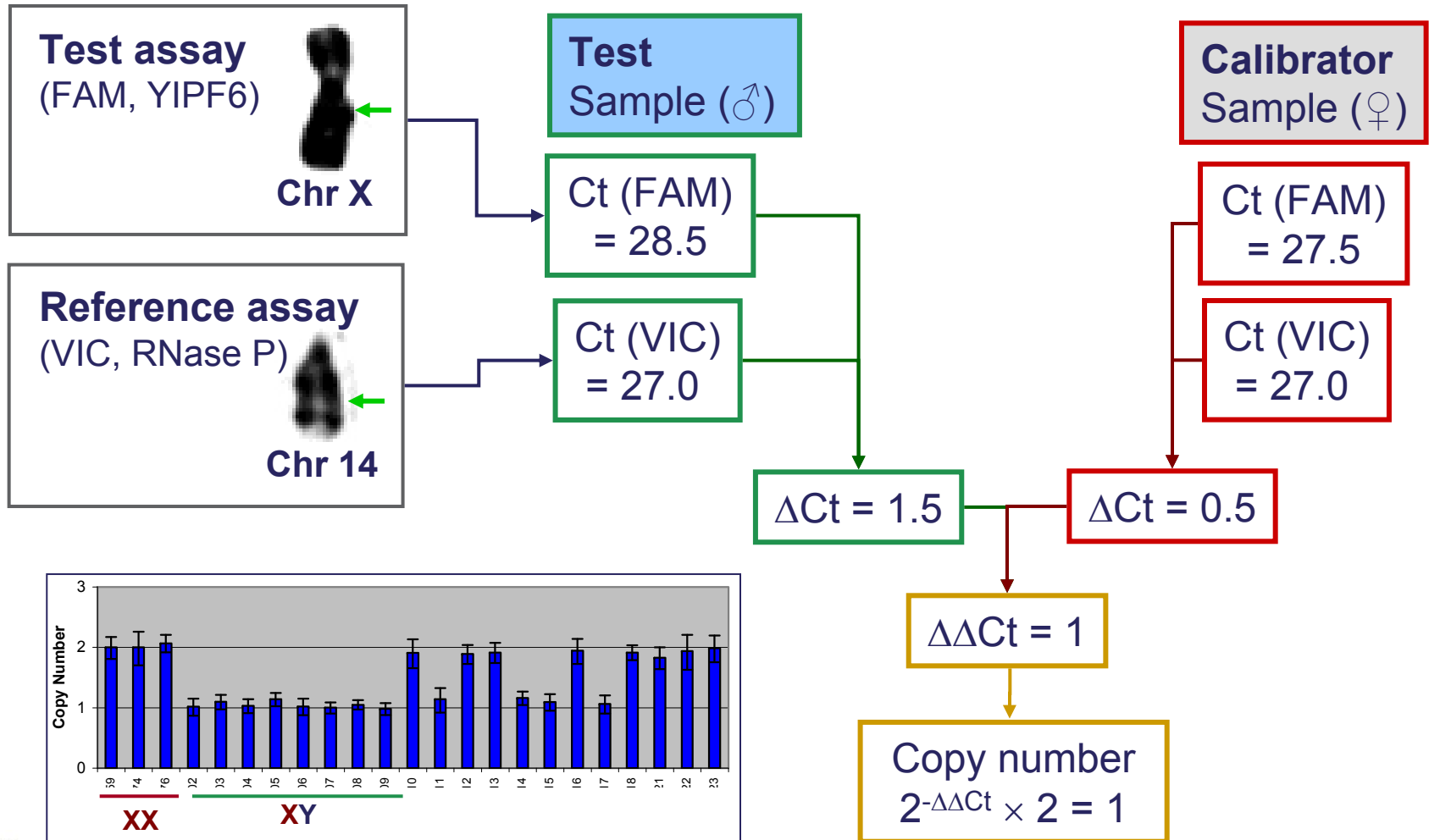
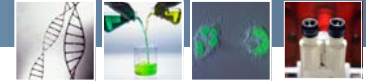
- quantitative detection of DNA copy number
- real-time PCR assays
- TaqMan-based
- duplex format
- simple workflow
- high sample throughput
- genome-wide pre-designed assay collection



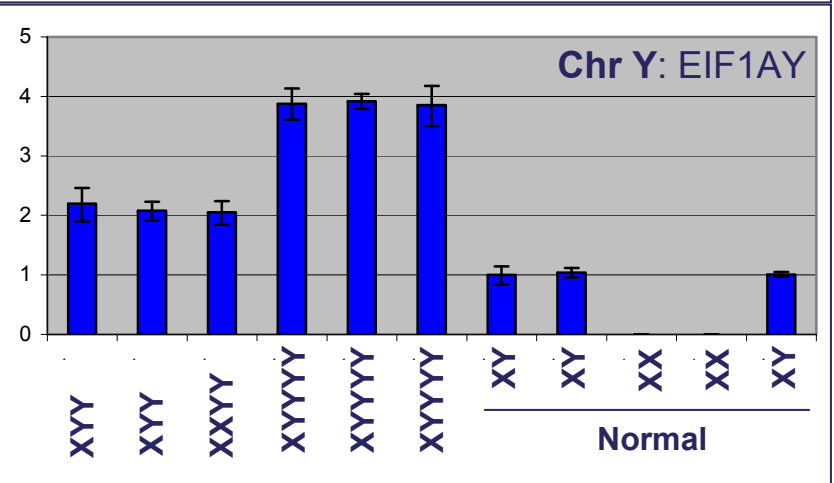
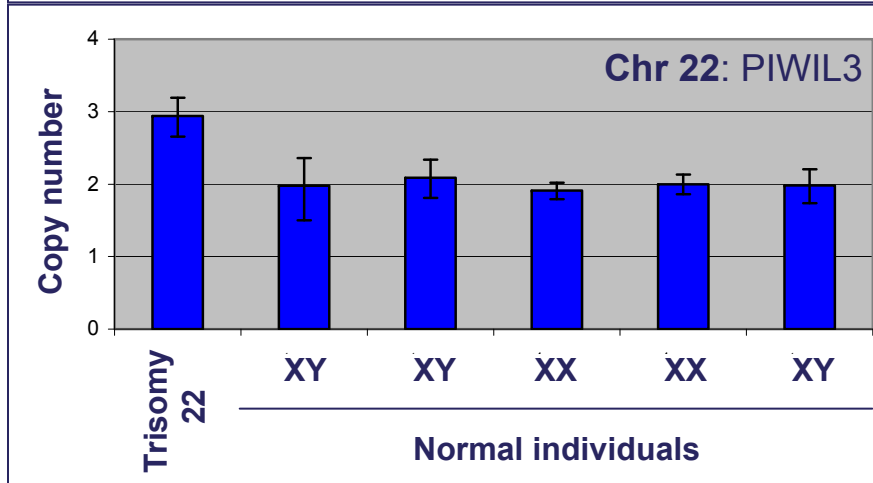
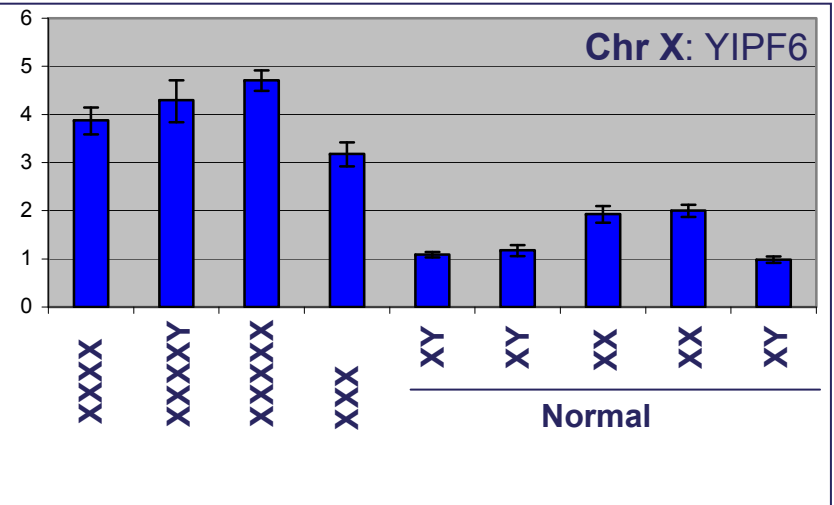
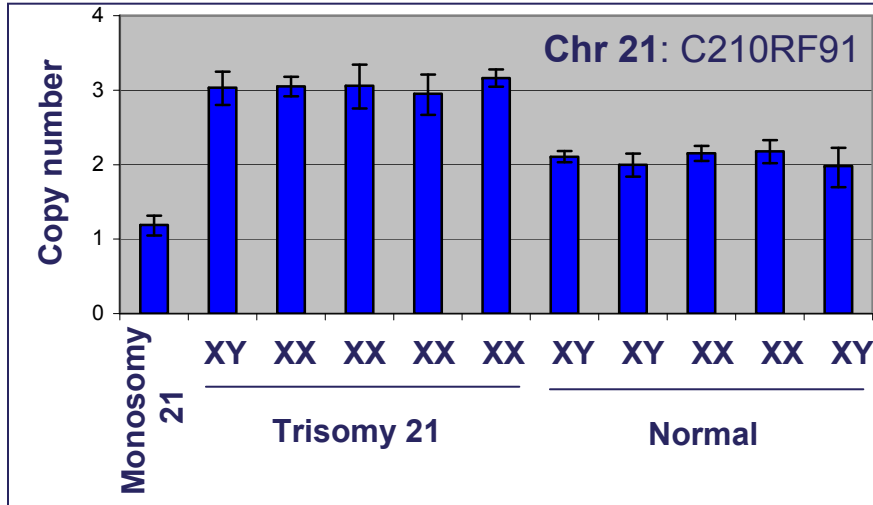
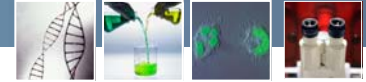
TaqMan chemistry



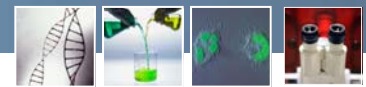
How Does Copy Number Assay Work?



Accuracy Demonstration with Aneuploidy Samples

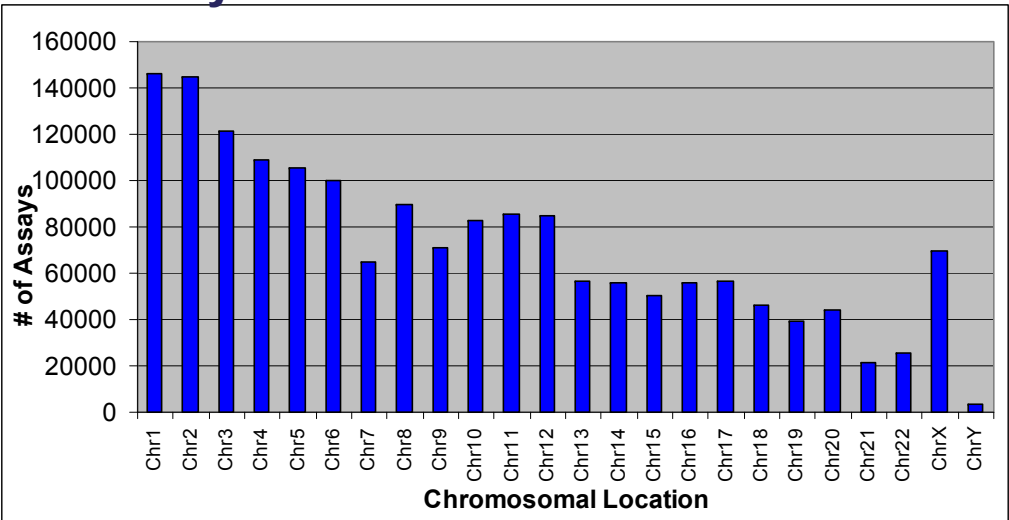


Pre-designed TaqMan Copy Number Assays - Genome Wide Coverage

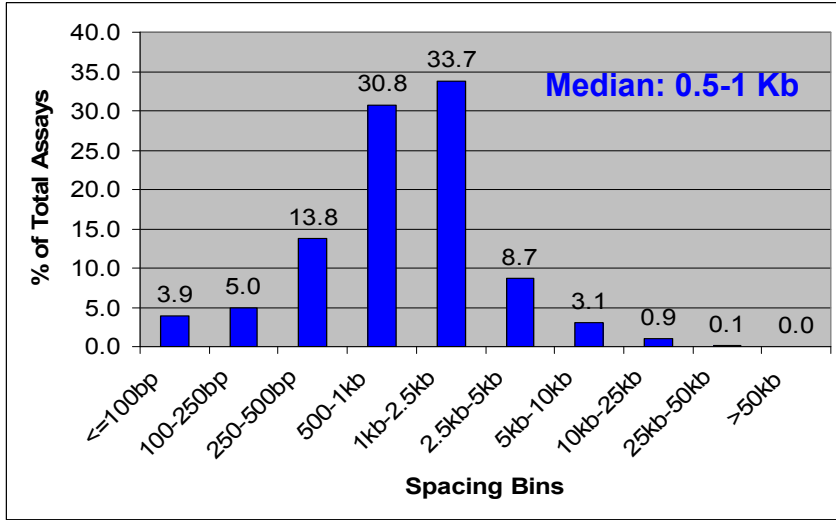


total # of assays	1,731,619
average # of assays for each gene or CNV	
40.8 assays per gene	
79.5 assays per CNV locus	

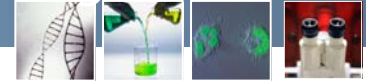
Assay Distribution on Chromosomes



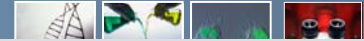
Assay Spacing Distribution



TaqMan® Assay Validation for CNV Discovery



- **Collaboration with Structural Genomic Variation Consortium**
 - Charles Lee, Harvard Medical School, Boston, MA, USA
 - Stephen Scherer, Sick Kids Hospital, Toronto, Canada
 - Nigel Carter and Matthew Hurles, Sanger Center, UK
- **High-resolution Copy Number Variation (CNV) Discovery** from CNV Consortium
 - Initially announced in July, 2007 by the Consortium
 - To develop a comprehensive, higher resolution CNV map for the Human Genome
 - at a level 100-fold finer than the first map (resolution 50Kb → 500bp)
 - 42,000,000 oligo probe sets on 20 x 2.1M Nimblegen arrays run over 40 individuals
- **Objective of the study**
 - To validate CNVRs that were identified by a high-resolution CNV discovery project of the Consortium



ARTICLES

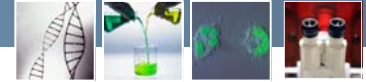
Nature advance online publication 7 October 2009

Origins and functional impact of copy number variation in the human genome

Donald F. Conrad^{1*}, Dalila Pinto^{2*}, Richard Redon^{1,3}, Lars Feuk^{2,4}, Omer Gokcumen⁵, Yujun Zhang¹, Jan Aerts¹, T. Daniel Andrews¹, Chris Barnes¹, Peter Campbell¹, Tomas Fitzgerald¹, Min Hu¹, Chun Hwa Ihm⁵, Kati Kristiansson¹, Daniel G. MacArthur¹, Jeffrey R. MacDonald², Ifejinelo Onyiah¹, Andy Wing Chun Pang², Sam Robson¹, Kathy Stirrups¹, Armand Valsesia¹, Klaudia Walter¹, John Wei², Wellcome Trust Case Control Consortium†, Chris Tyler-Smith¹, Nigel P. Carter¹, Charles Lee⁵, Stephen W. Scherer^{2,6} & Matthew E. Hurles¹

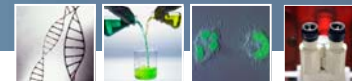
Structural variations of DNA greater than 1 kilobase in size account for most bases that vary among human genomes, but are still relatively under-ascertained. Here we use tiling oligonucleotide microarrays, comprising 42 million probes, to generate a comprehensive map of 11,700 copy number variations (CNVs) greater than 443 base pairs, of which most (8,599) have been validated independently. For 4,978 of these CNVs, we generated reference genotypes from 450 individuals of European, African or East Asian ancestry. The predominant mutational mechanisms differ among CNV size classes. Retrotransposition has duplicated and inserted some coding and non-coding DNA segments randomly around the genome. Furthermore, by correlation with known trait-associated single nucleotide polymorphisms (SNPs), we identified 30 loci with CNVs that are candidates for influencing disease susceptibility. Despite this, having assessed the completeness of our map and the patterns of linkage disequilibrium between CNVs and SNPs, we conclude that, for complex traits, the heritability void left by genome-wide association studies will not be accounted for by common CNVs.

The CNV Validation Study Design



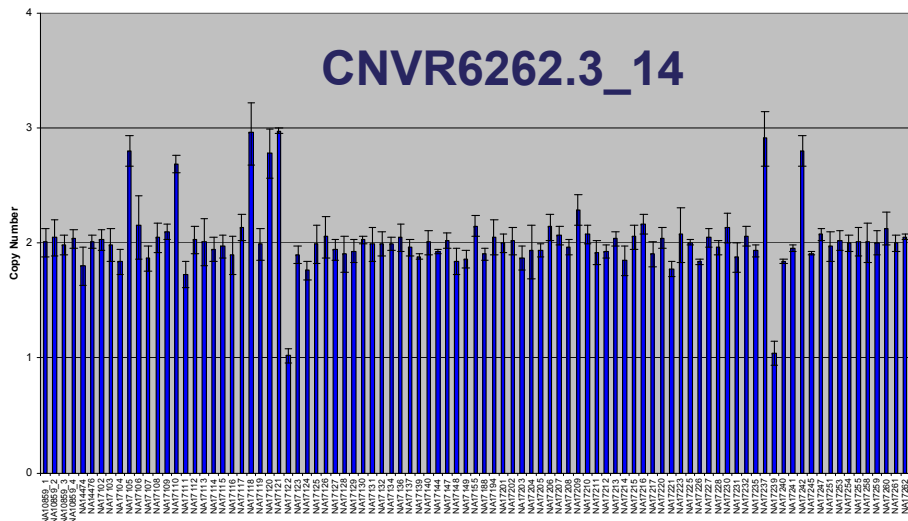
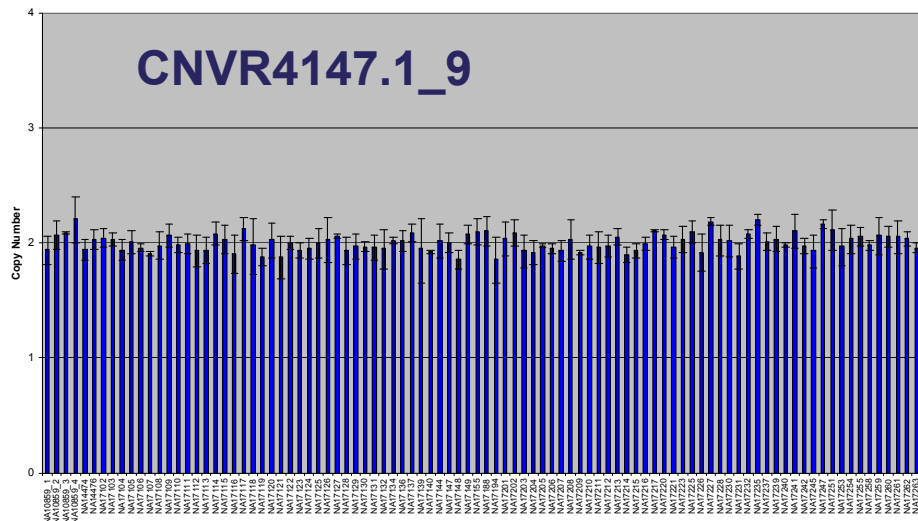
- About 100 CNVRs are selected for TaqMan assay validation
- One assay per CNVR for most targets, as to cover as many CNVRs
- Roughly half of selected targets are in gene regions.
- Pre-designed TaqMan copy number assays are randomly selected
- These assays were validated first with a panel of 92 gDNA samples
- The 40 HAPMAP samples (20 CEU and 20 YRI) and the reference sample NA10851 were tested with selected assays.
- TaqMan and array data comparison are analyzed

Examples of Copy Number Profiles



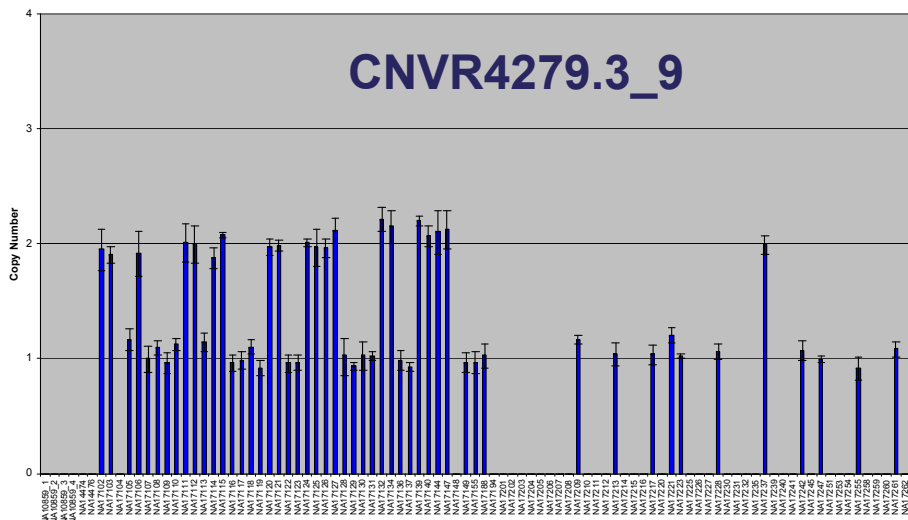
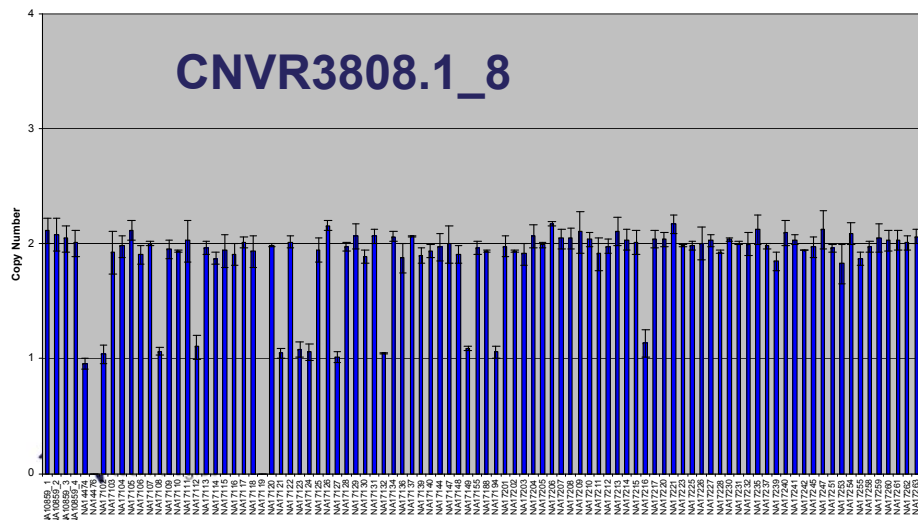
TaqMan CNV assay
7900 - 10 uL - 10 ng dry gDNA
20081217_CNV4147.1_9_RNaseP_4_newDME_Yojimbo_IC_SDS23

TaqMan CNV assay
7900 - 10 uL - 10 ng dry gDNA
20081218_CNVR6262.3_14_RNaseP_4_newDME_Yojimbo_BC_SDS23

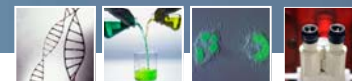


TaqMan CNV assay
7900 - 10 uL - 10 ng dry gDNA
20081217_CNVR3808.1_8_RNaseP_4_newDME_Yojimbo_IC_SDS23

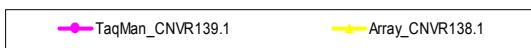
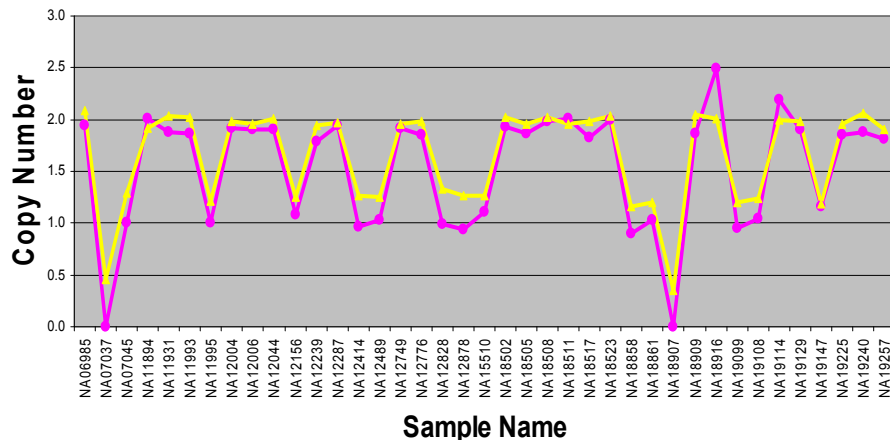
TaqMan CNV assay
7900 - 10 uL - 10 ng dry gDNA
20081217_CNVR4279.3_9_RNaseP_4_newDME_Yojimbo_IC_SDS23



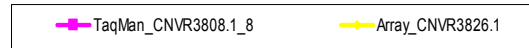
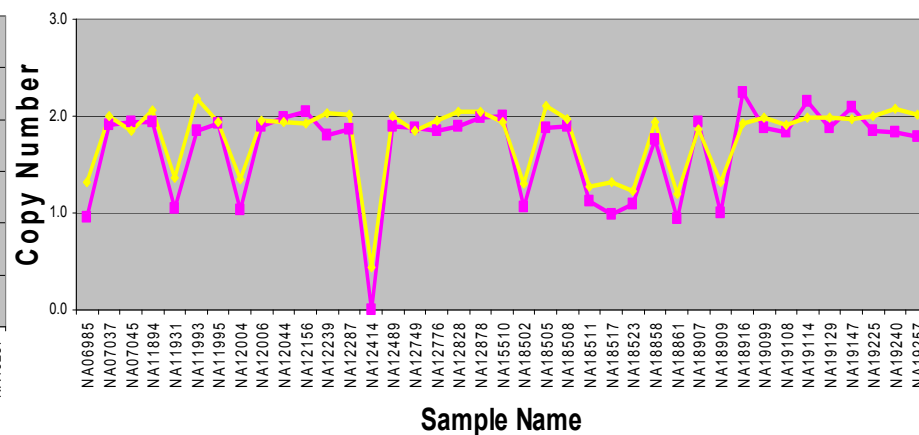
Examples of Good Correlation -- with higher frequency of CN changes



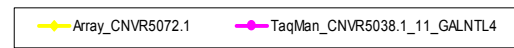
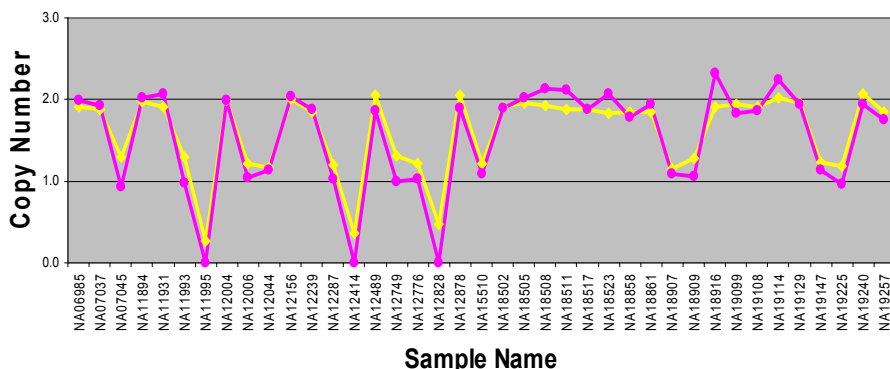
CNVR138.1



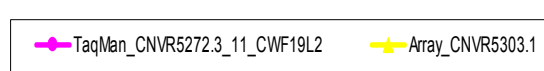
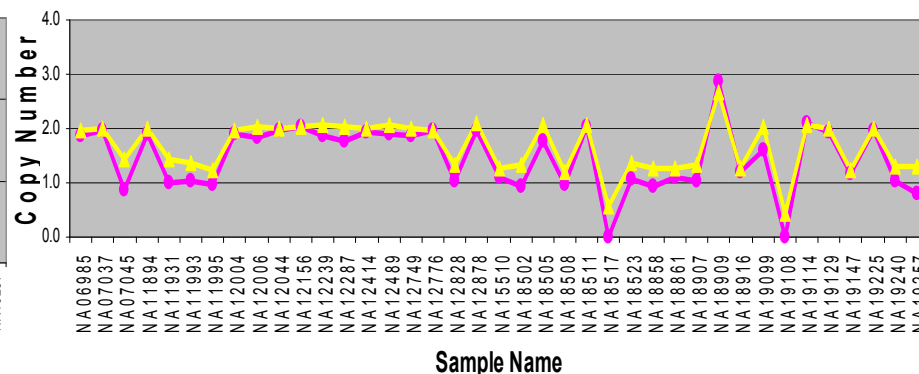
CNVR3826.1



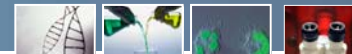
CNVR5072.1



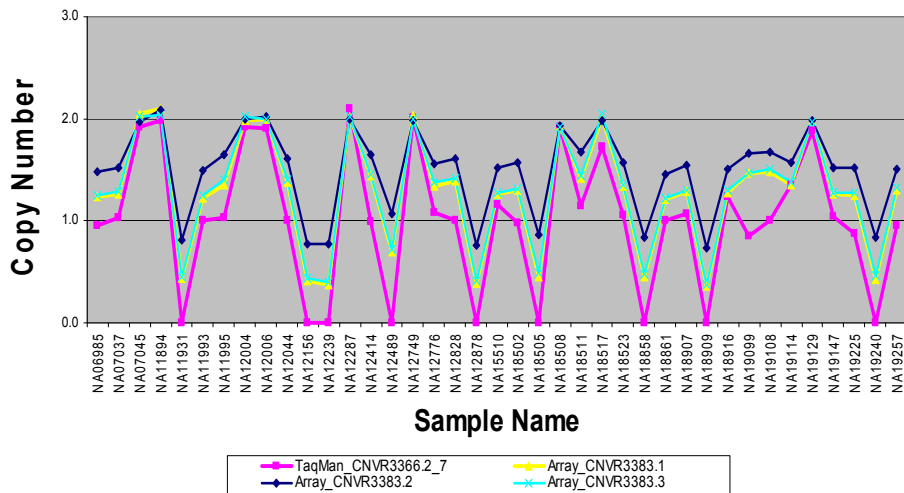
CNVR5303.1



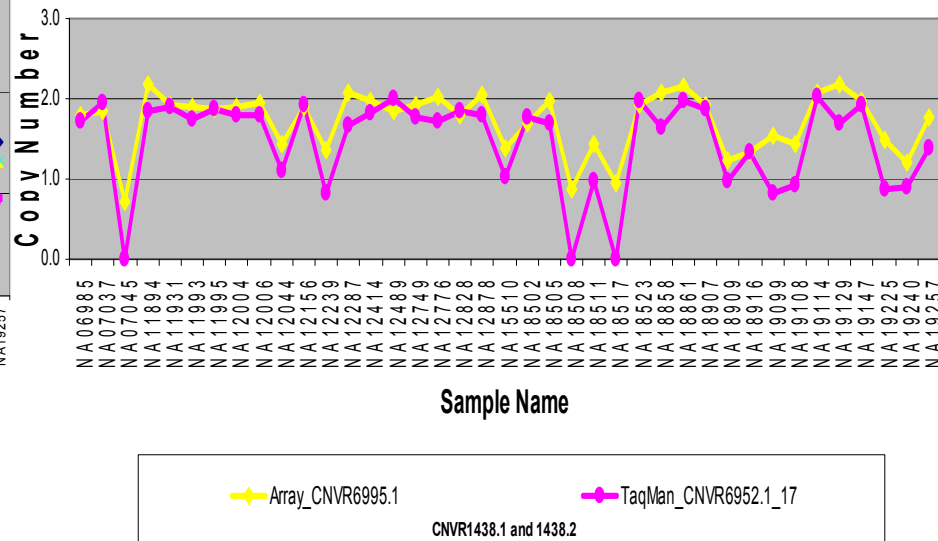
Good correlation, but array data have narrower “dynamic range”



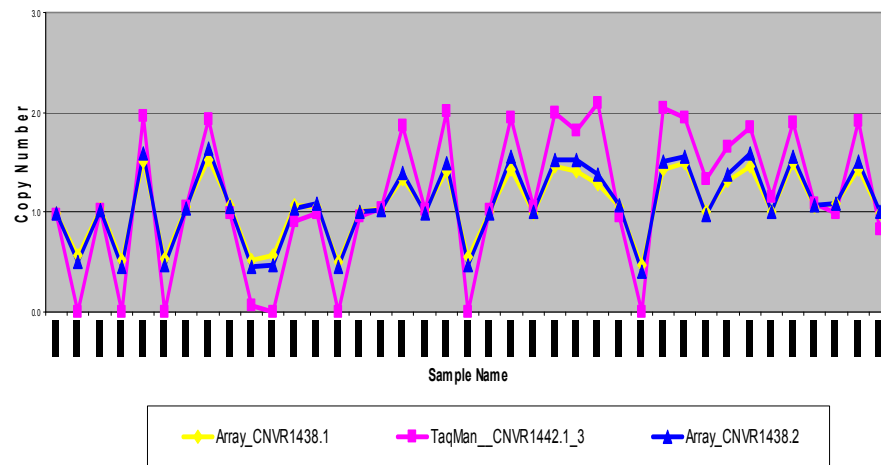
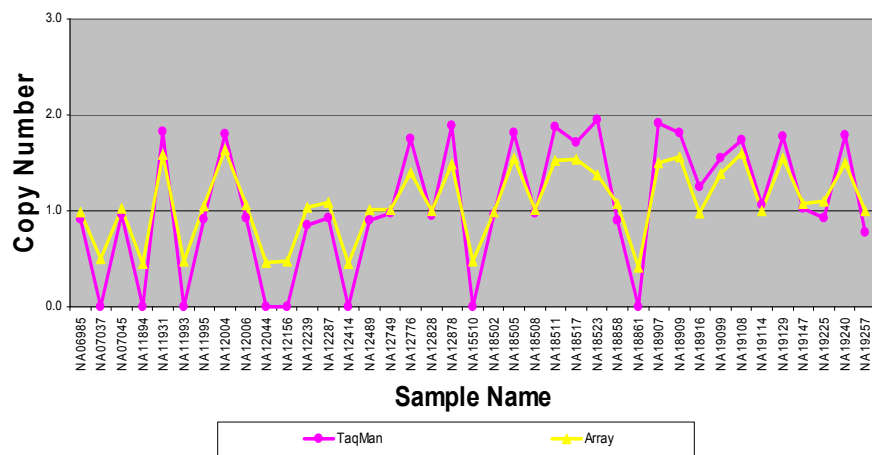
CNVR3383.1



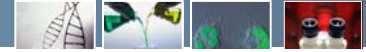
CNVR6995.1



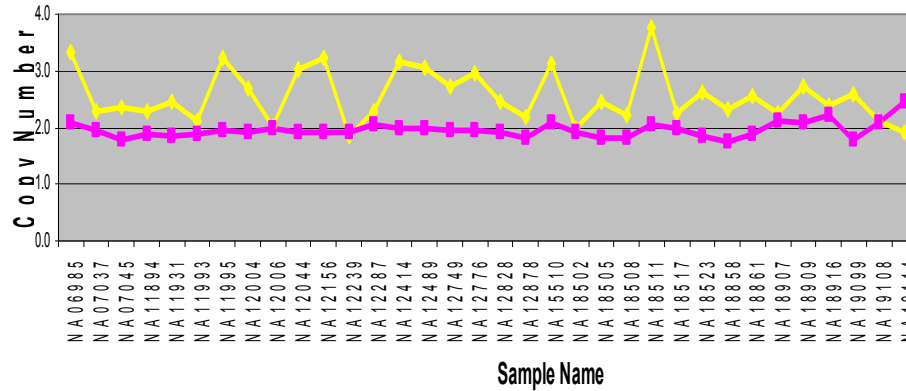
CNVR1438.1



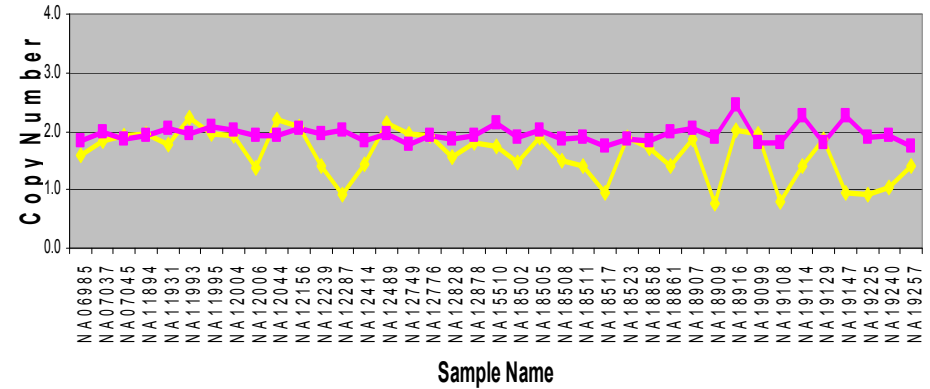
False positives in array data



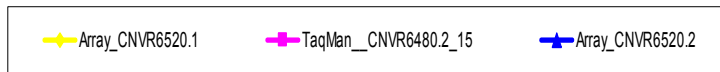
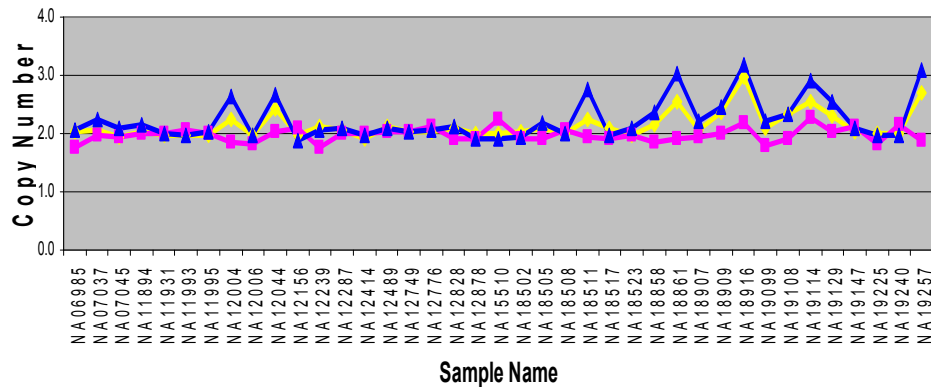
CNVR7471.1



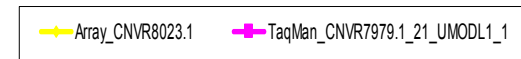
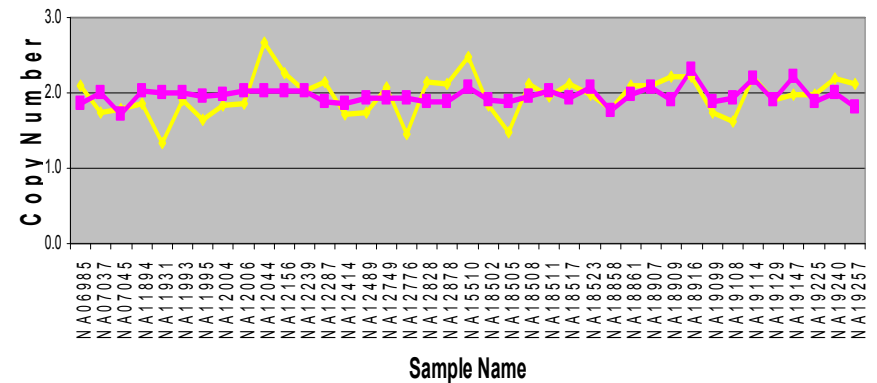
CNVR5036.1



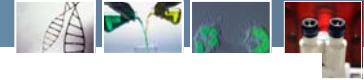
CNVR6520.1 and 6520.2



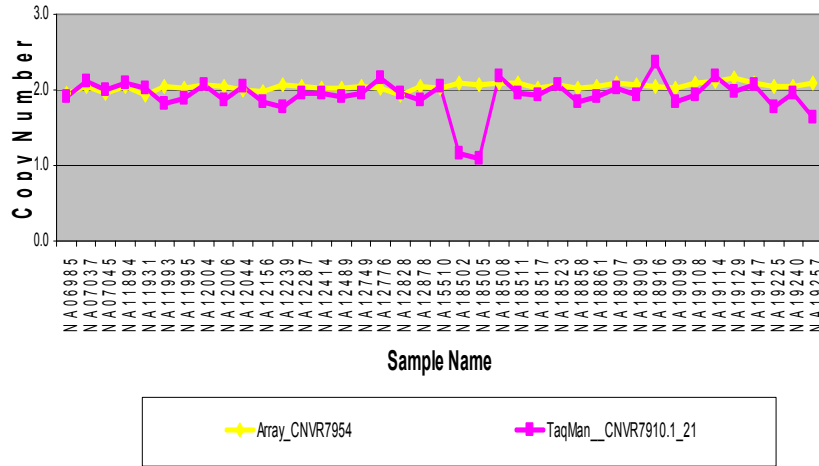
CNVR8023.1



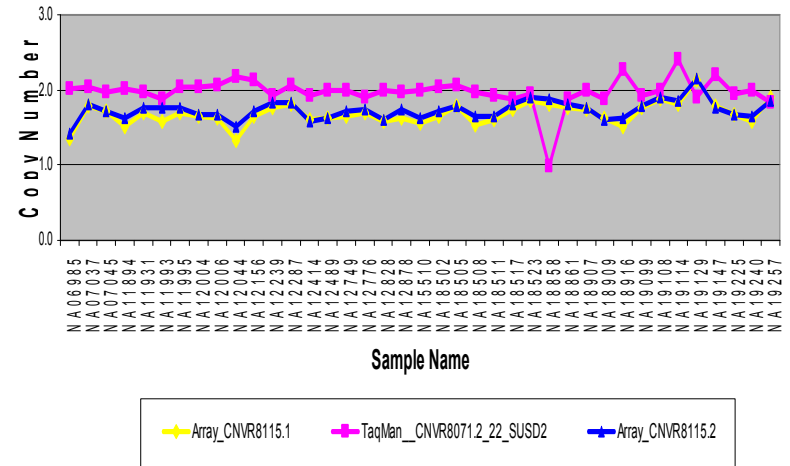
False negatives in array data



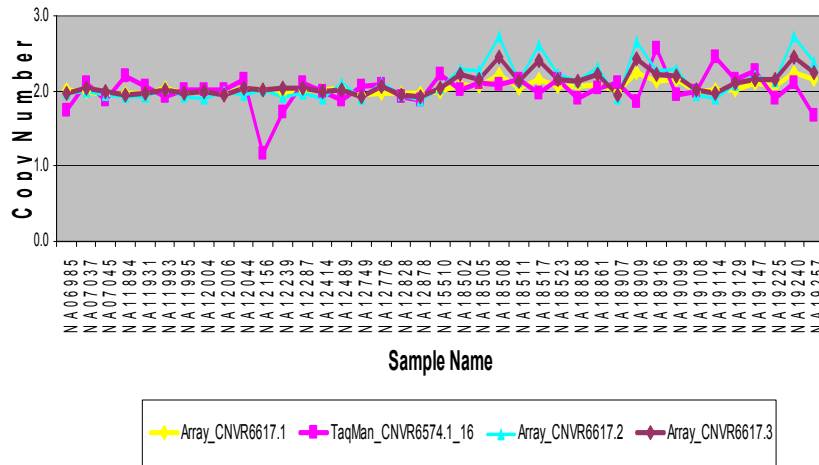
CNVR7954.1



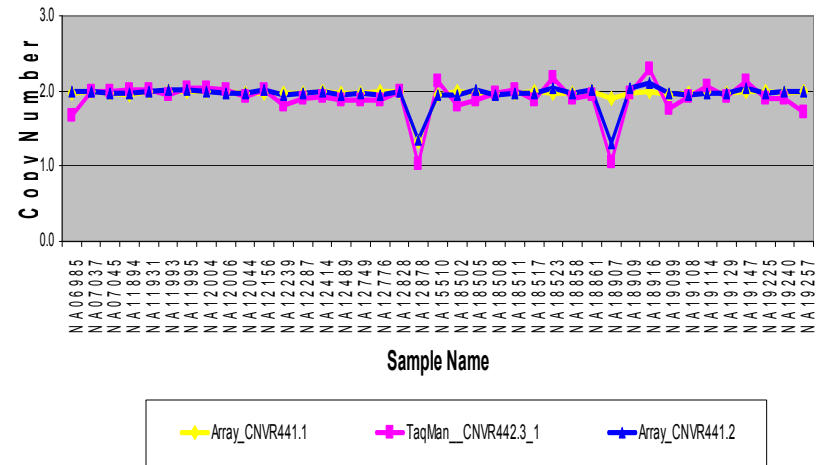
CNVR8115.1 and 8115.2



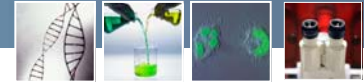
CNVR6617.1, 6617.2 and 6617.3



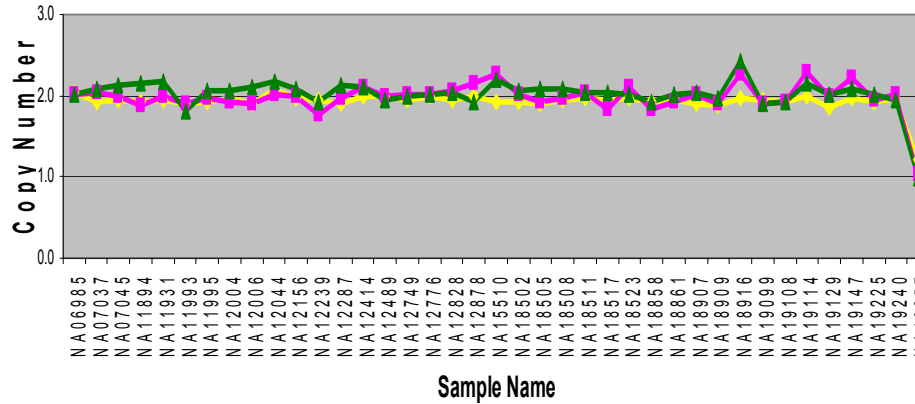
CNVR441.1 and 441.2



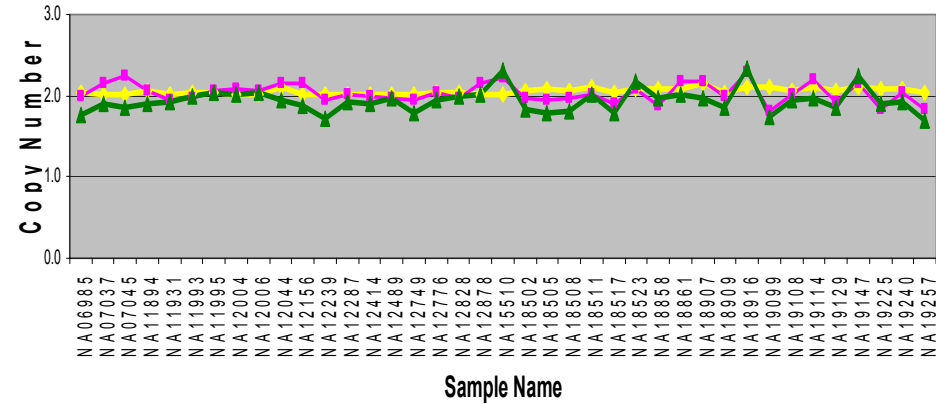
Two TaqMan assays targeting a CNVR



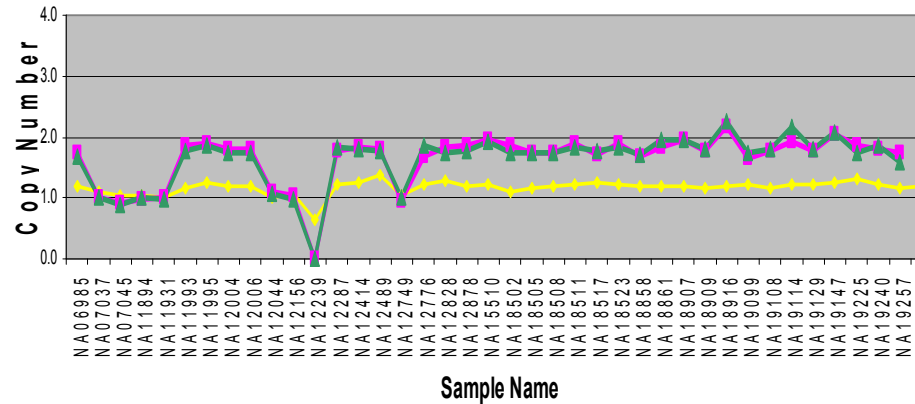
CNVR7811.1



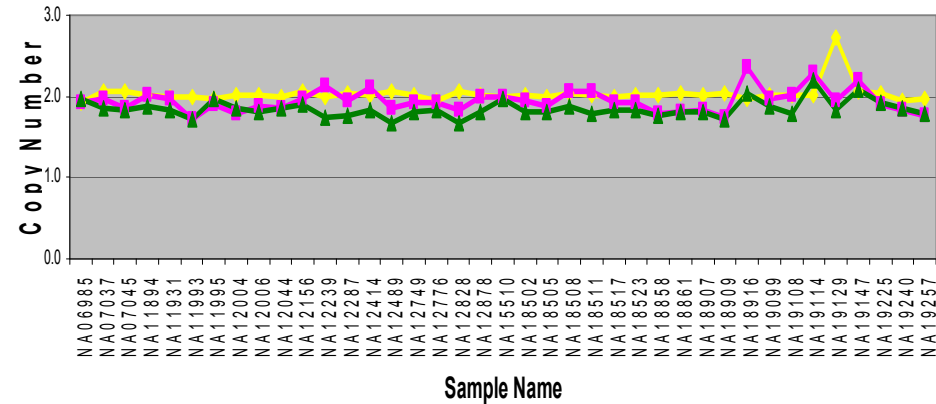
CNVR5450.1



CNVR7722.1_Ref sample CN=1

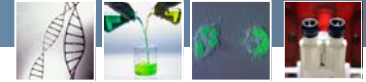


CNVR369.1



◆ Array_CNVR7722.1
 ■ TaqMan_CNVR7811.1_19
 ▲ TaqMan_CNVR7811.1_19_LILRA3_1

◆ Array_CNVR369.1
 ■ TaqMan_CNVR370.1_1_GBA_1
 ▲ TaqMan_CNVR370.1_1_MTX1



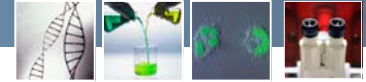
- **Objective**

- Identify deletion/duplication in 1000+ clinical samples by screening with TaqMan Copy Number Assays in targeted regions for further breakpoint studies
- Demonstrate the utility of TaqMan copy number assay in studies with large sample screening

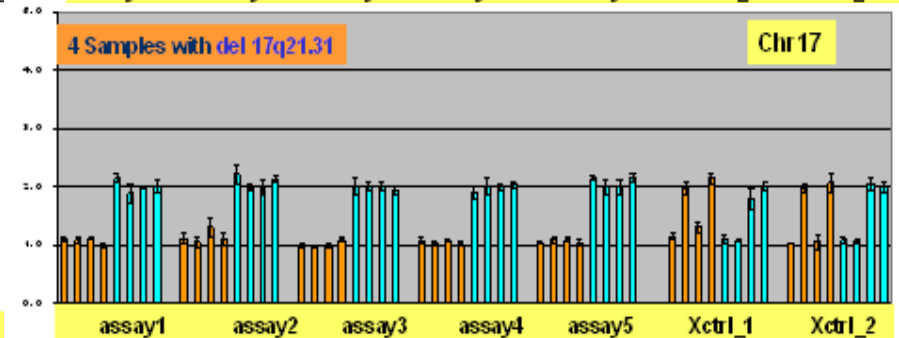
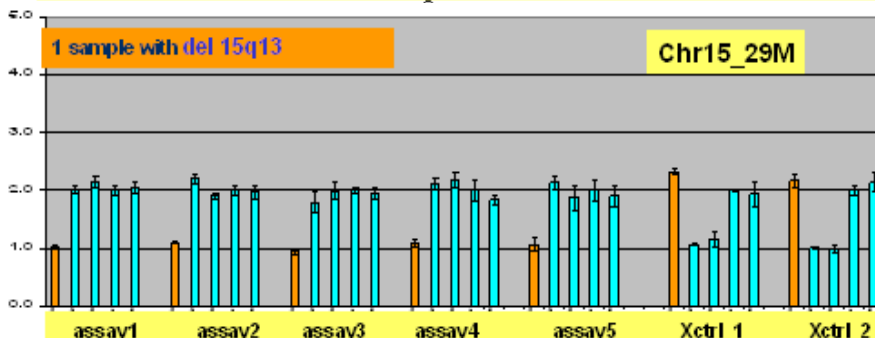
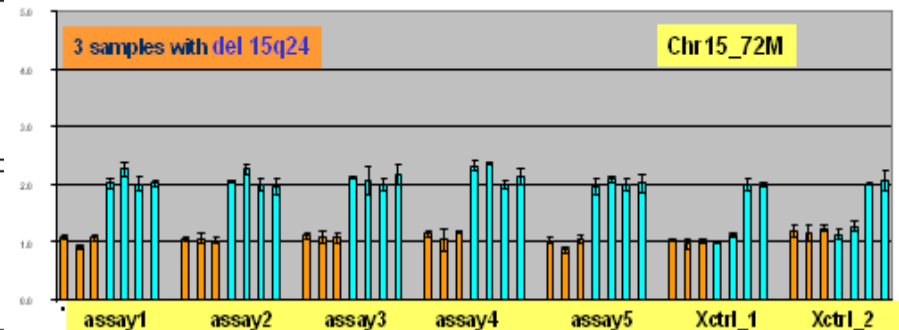
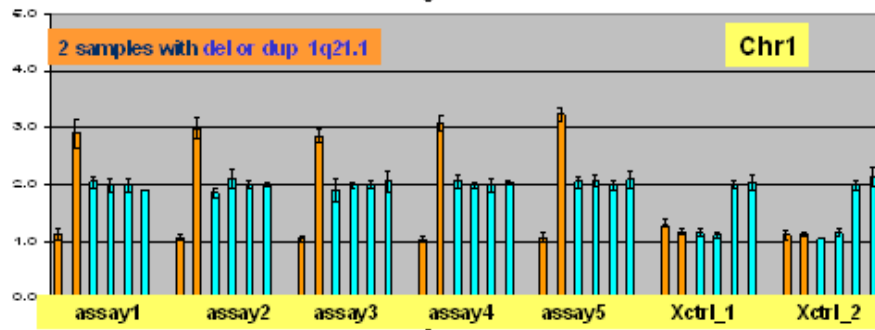
- **Design**

- 4 targets from 4 chromosomal regions and 5 assays/region
- “Control” assays from X chromosome
- These assays were validated with an internal panel first
- Positive clinical samples were provided and tested with the validated assays
- 1000+ clinical samples were received and tested

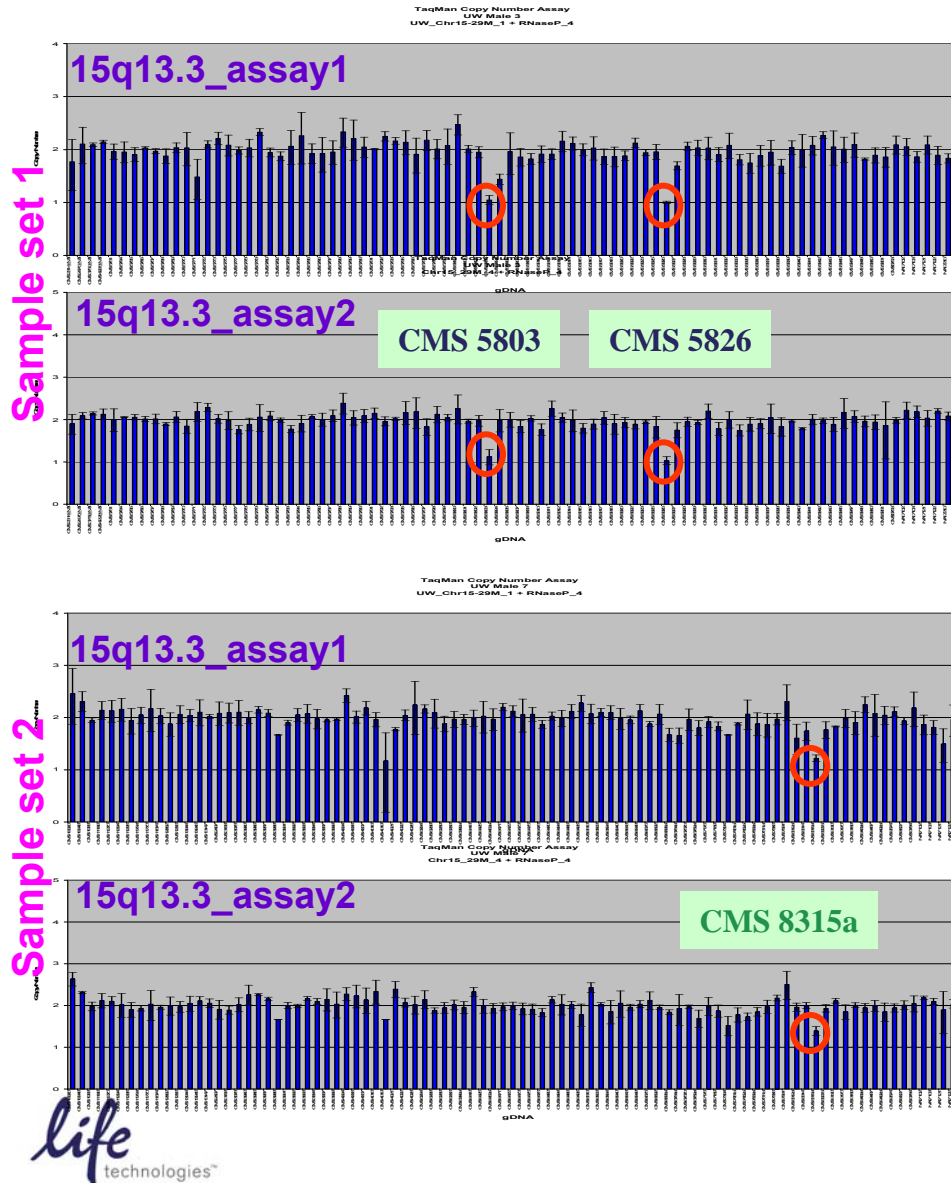
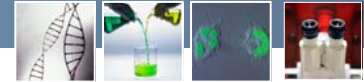
Targeted Regions and Their Assay Validation



chromosomes and coordinates	size of the regions
chr1:145100000-145800000	700Kb
chr15:29000000-29500000	500Kb
chr15:72300000-73500000	1.2Mb
chr17:41100000-41500000	500Kb



Screening 1000+ Clinical Samples



nature
genetics

VOLUME 40 | NUMBER 3 | MARCH 2008 NATURE GENETICS

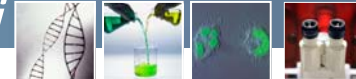
A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures

The NEW ENGLAND JOURNAL of MEDICINE

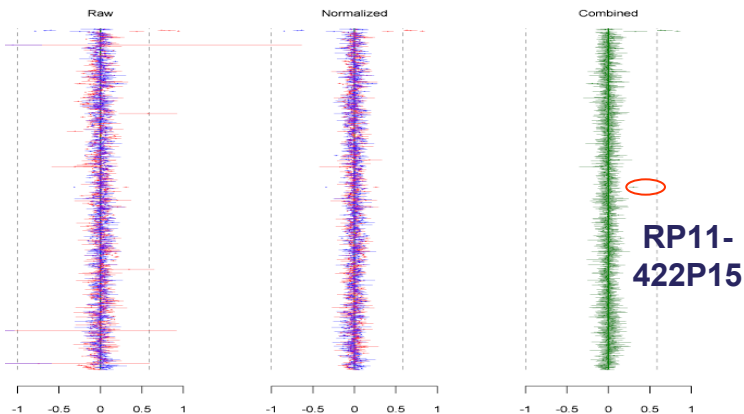
Recurrent Rearrangements of Chromosome 1q21.1 and Variable Pediatric Phenotypes

N Engl J Med 359:1685, October 16, 2008

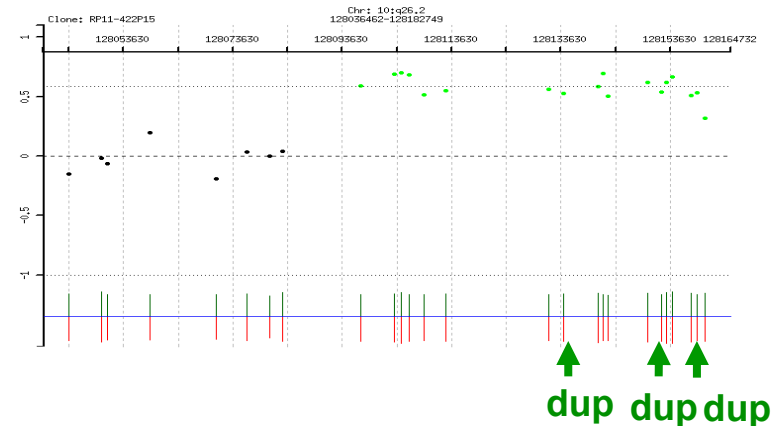
Clarification of Copy Number Ambiguity from Other platforms – Collaboration with Dr. Bi



BAC Array CGH



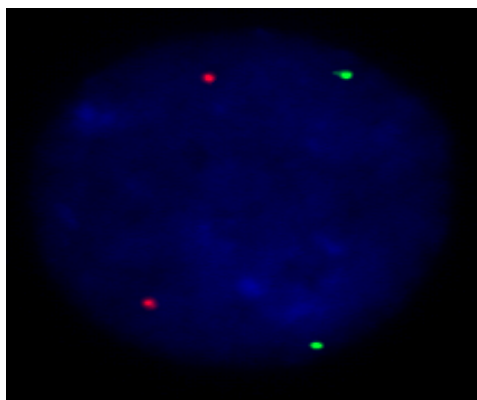
Oligo Array CGH



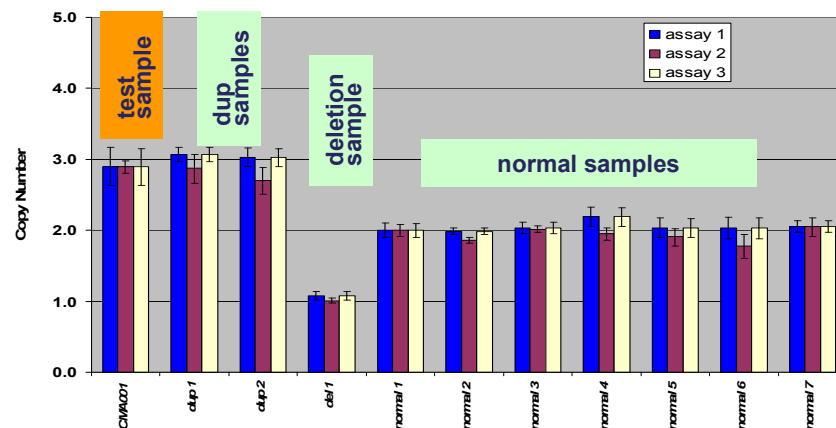
FISH

Red:
RP11-422P15

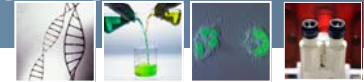
Green:
a control probe
in Chr 10q



TaqMan Copy Number Assay

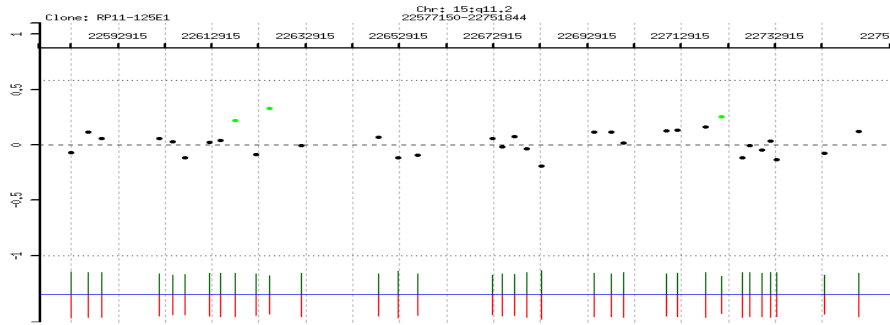


Concordance between TaqMan Assays and Oligo aCGH



RP11-125E

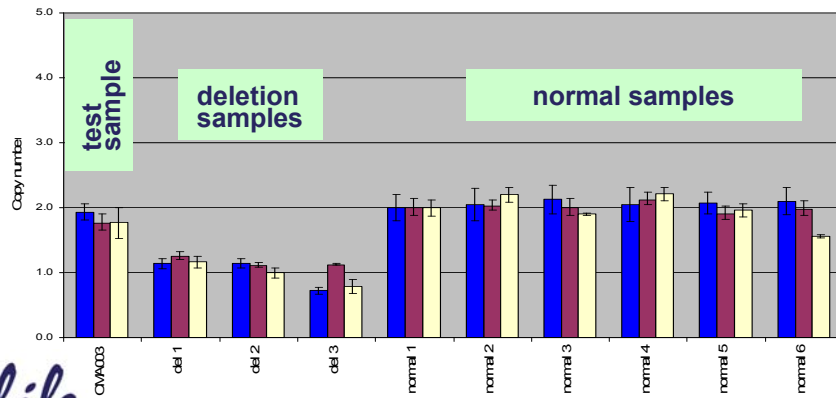
Oligo Array CGH



TaqMan Assay Location

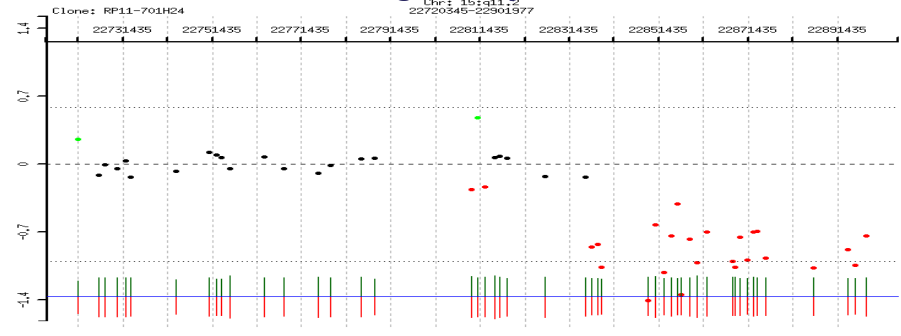
↑ normal normal normal

TaqMan Copy Number Assay



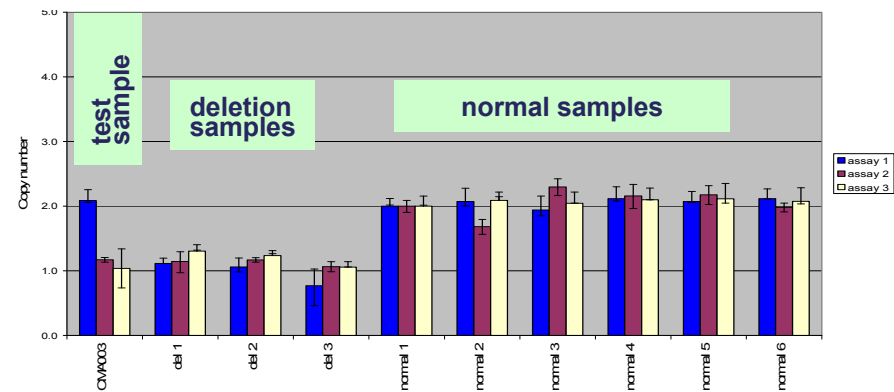
RP11-701H24

Oligo Array CGH

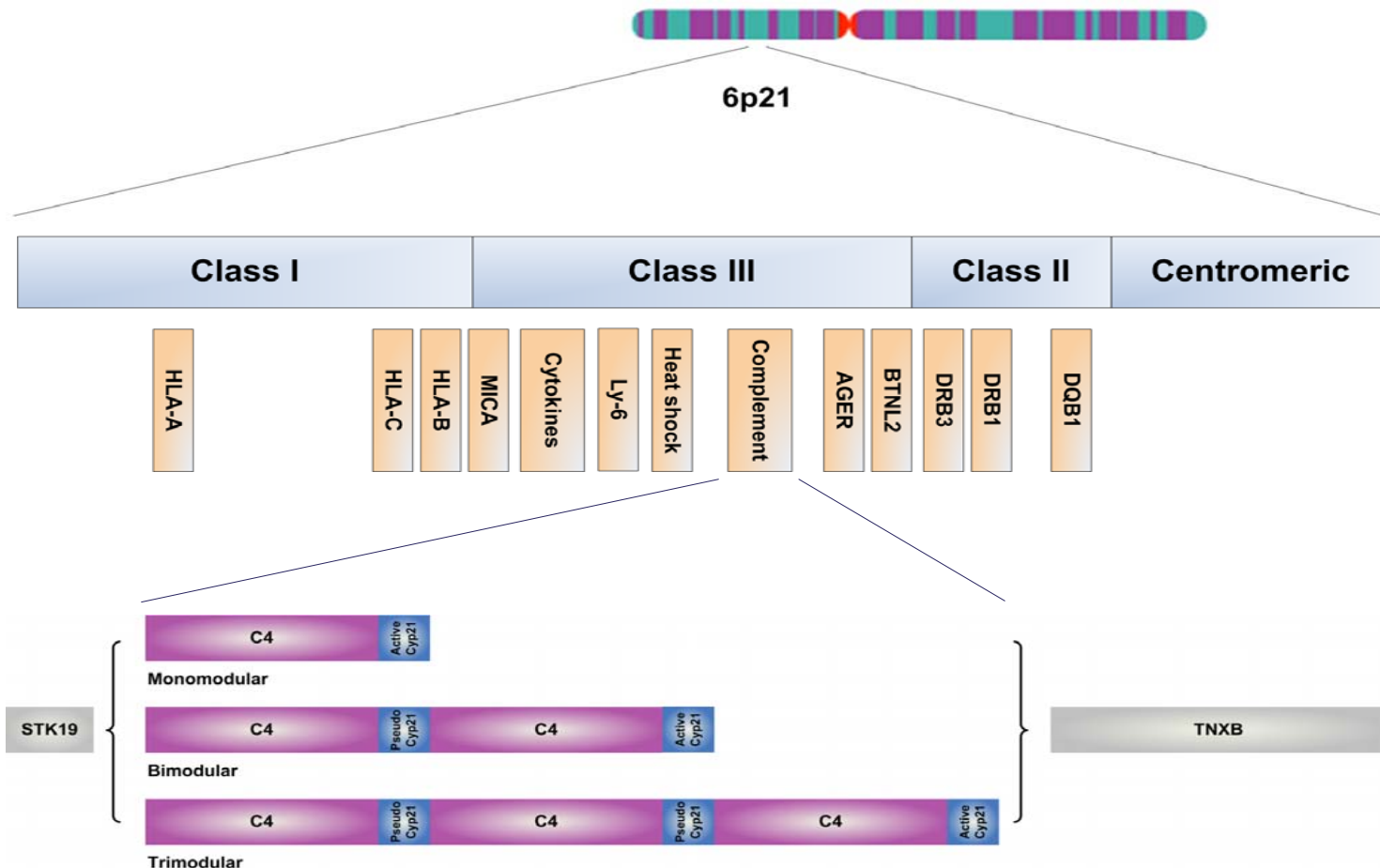
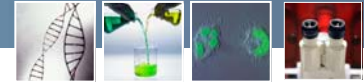


↑ normal ↑ del del

TaqMan Copy Number Assay



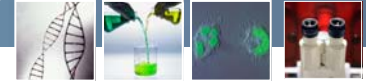
The Complement Factor 4 (C4) Region



• The region is associated with several autoimmune diseases

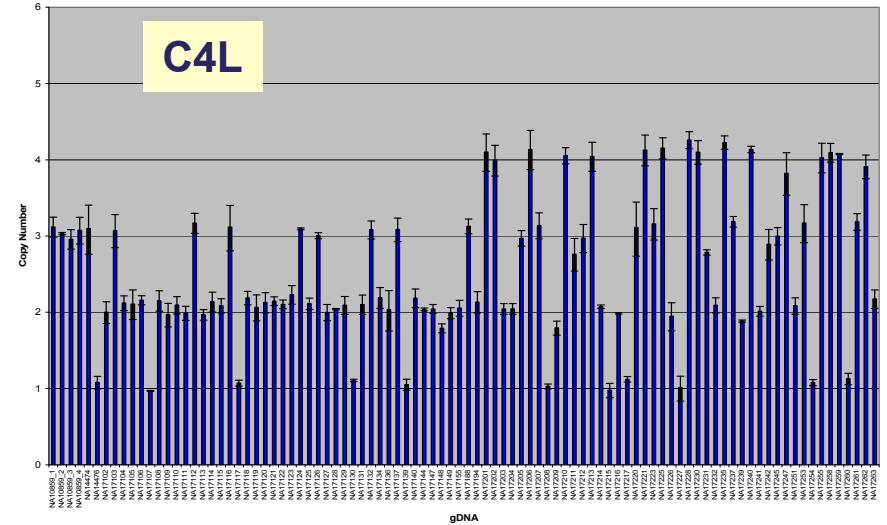
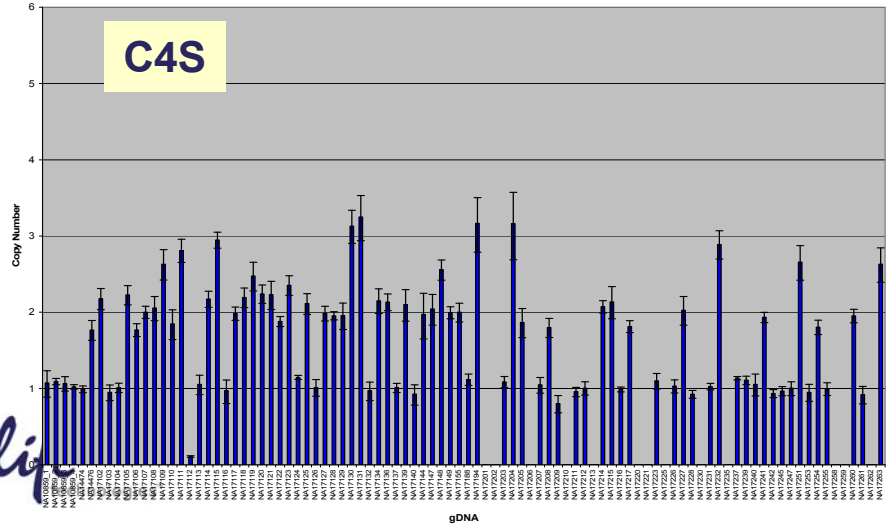
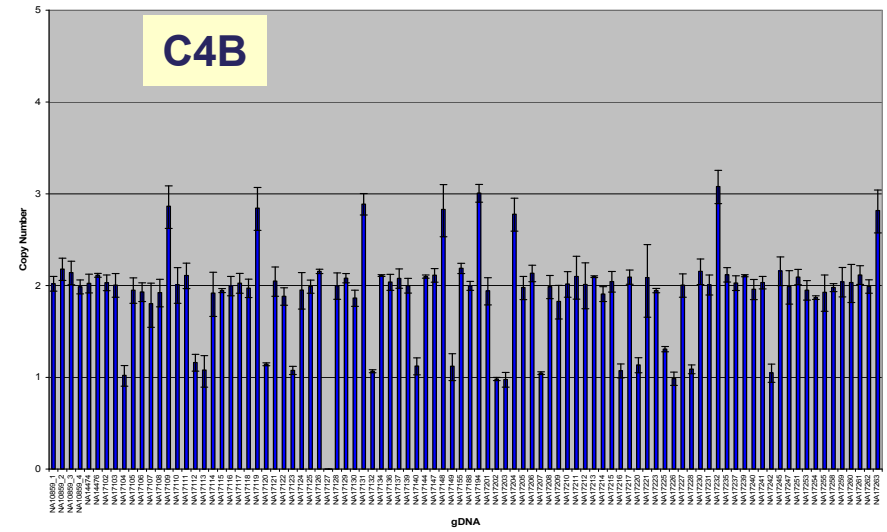
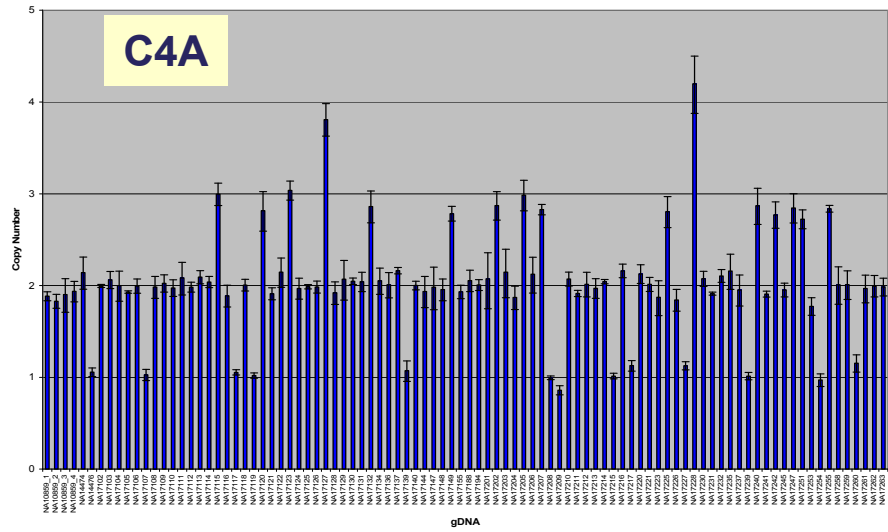
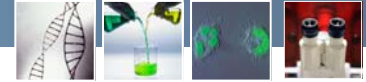
• Low copy number of C4A increases risk of SLE. Yang et al. (Am J Hum Genet, 2007)

Complexity of Structural Variation at C4 Locus



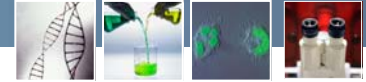
- Number of genes (C4 CN = 0-8)
- Functional variants (C4a/C4b – 5 SNPs)
 - C4a: CCTGTCCAGTGTTAGAC
 - C4b: TCTCTCCAGTGATACAT
 - Long half-life: C4a protein variant
 - Short half-life: C4b protein variant
- InDel within C4 gene:
 - endogenous retrovirus (HERV-C4)
 - 6 kb insertion in intron 9
 - presence: long C4 gene (C4L)
 - absence: short C4 gene (C4S)
 - unknown function (promoter activity?)
- *Challenge: determine allele-specific CN of C4*

Isoform-specific Copy Number Detection for C4 Locus



lif

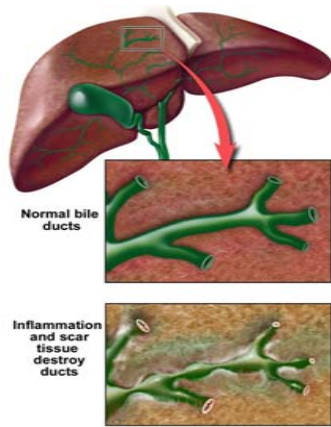
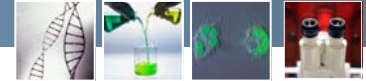
Validation of C4 Assays



- Isoform-specific TaqMan Copy Number Assays for C4A, C4B, C4S and C4L were designed and tested
- These assays were validated with the 10 samples that have known C4 copy number for each isoform

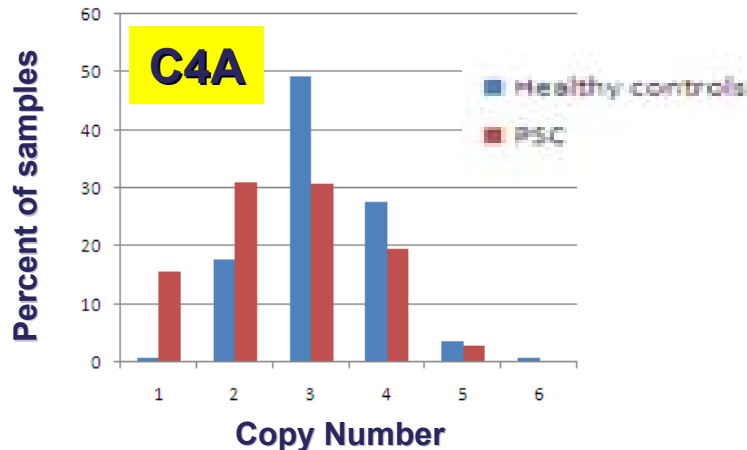
Cell line	C4A	C4B	C4 long (HERV present)	C4 short (HERV absent)	ABI TaqMan	Sequencing (Wellcome)	Southern blot/RFLP validated Real-Time PCR (Ohio)
IHWG9020	2	0	2	0	YES	YES	YES
IHWG9022	0	2	0	2	YES	YES	YES
IHWG9029	4	0	4	0	YES	N/A	YES
IHWG9065	2	2	4	0	YES	N/A	YES
IHWG9098	2	2	4	0	YES	N/A	YES
IHWG9048	2	2	2	2	YES	N/A	YES
IHWG9061	4	2	6	0	YES	N/A	YES
IHWG9060	4	4	6	2	YES	N/A	YES
IHWG9005	3	2	4	1	YES	N/A	YES
IHWG9017	1	2	3	0	YES	N/A	YES

C4A Copy Number Association with Primary Sclerosing Cholangitis (PSC)

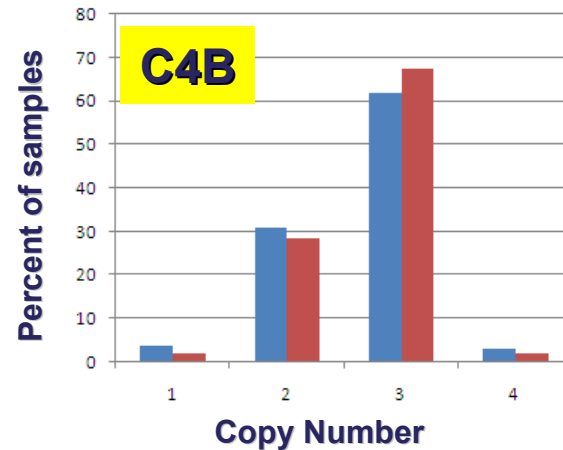


- chronic inflammation and fibrosis lead to intra- and extrahepatic bile duct strictures and eventually liver failure
- due to prevalence ~ 1/10.000
- ↑ autoimmune diseases
- ↑ inflammatory bowel disease
- ↑ cancer

With 251 patients and 253 healthy controls



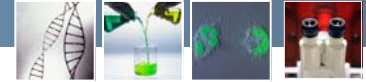
Strong disease association



No disease association

Exon-specific Copy Number Detection for OTC Gene

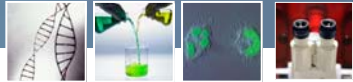
-- Collaboration with Dr. Allen Bale



X- Linked Ornithine Transcarbamylase Deficiency

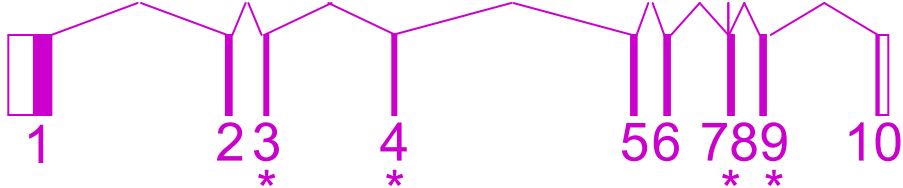
- OTC is an important enzyme in the urea cycle
- OTC gene, located at Xp21.1, spans 73 kb with an ORF of 1,062 nucleotides distributed into 10 exons
- Defects in the OTC gene cause a block in ureagenesis resulting in hyperammonemia, which can lead to brain damage and death
- Approximately 10-15% of OTC deficiency is caused by gross deletions of one or more exons of the OTC gene
- Currently Southern blotting is used in its deletion detection in clinical labs

Design of Copy Number Assays for Various OTC Deletions



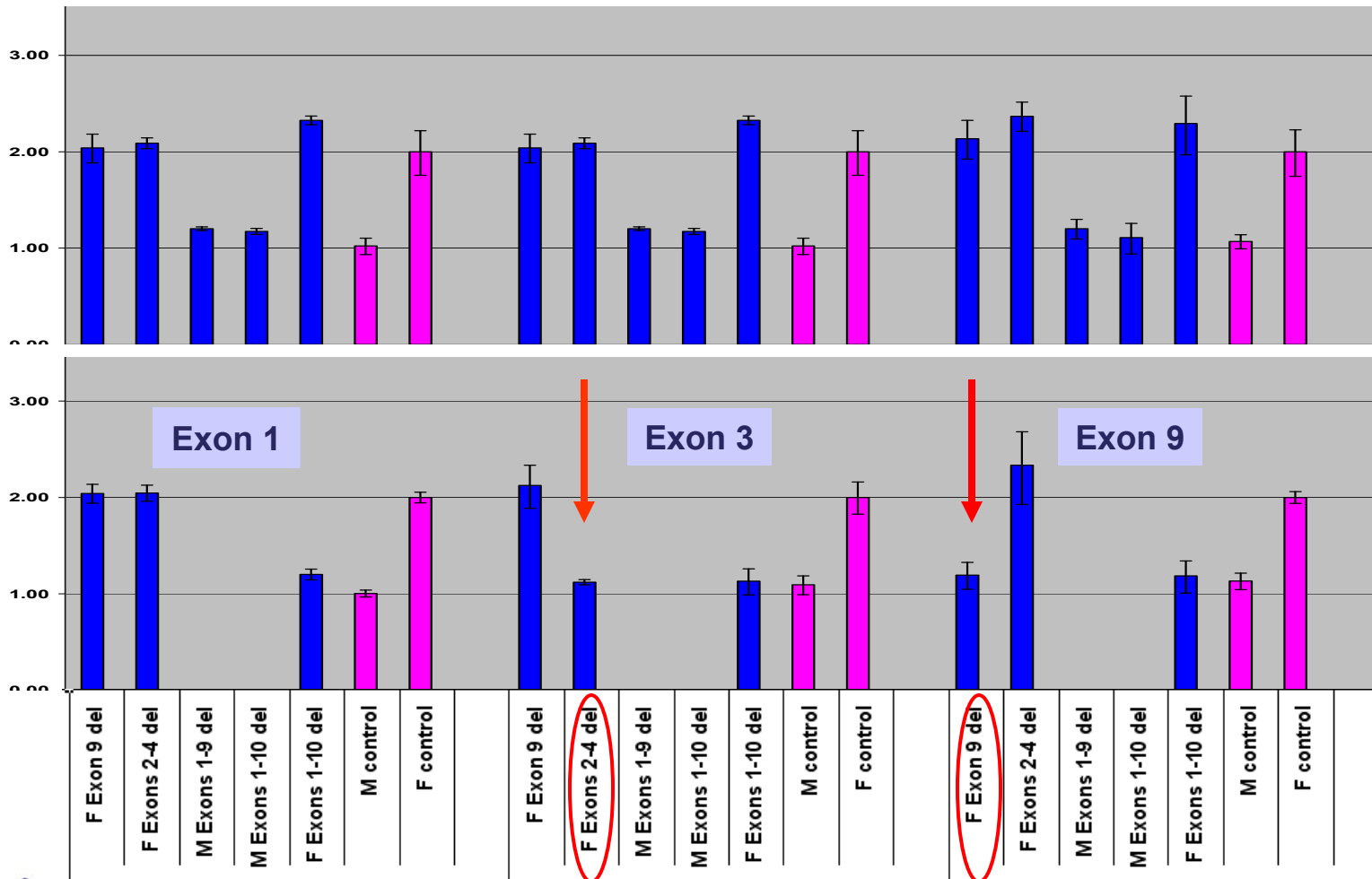
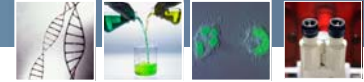
Types of Deletion

- Entire gene *
- Exons 1-3
- Exons 1-9 *
- Exons 1-8 *
- Exons 2-4 *
- Exon 3
- Exons 4-10
- Exons 7-9
- Exon 9 *



Types of Deletion	intron 1	intron 2 to intron 3	intron 8 to intron 9
		exon 3	exon 9
•Entire gene *	+	+	+
•Exons 1-3	+	+	
•Exons 1-9 *	+	+	+
•Exons 1-8 *	+	+	
•Exons 2-4 *		+	
•Exon 3		+	
•Exons 4-10			+
•Exons 7-9			+
•Exon 9 *			+

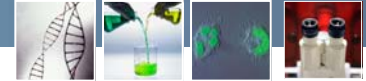
Exon-Specific OTC Deletion Detection by TaqMan Assays



X-ctrl assays

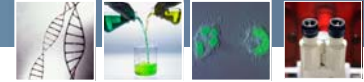
OTC assays

Summary



- CNV is a recently appreciated, widespread form of polymorphism in the human genome and can be associated with human diseases.
- TaqMan[®] copy number assays are designed to accurately detect copy number changes
 - provides genome-wide 1.7 million pre-designed assays as well as custom design options
 - Median assay spacing is between 0.5 -1 Kb
- Advantages of TaqMan[®] Copy Number Assays
 - Easy workflow, high sample throughput, and high specificity
 - High design and success rate
 - Highly quantitative and capable of determining the DNA copy number from 0 to 5 copies
- Applications
 - Validate CNV discovery from microarrays
 - Screen large samples for disease-associated CNVs or smaller InDels
 - Detect Isoform-specific copy number variation and exon-specific deletion
 - Confirm gene knockouts, transgenic animals

Acknowledgements of the Collaborators



➤ CNV Validation

Structural Genomic Variation Consortium

- **Dr. Charles Lee** and **Dr. Omer Gokcumen**, Harvard Medical School, Boston
- **Dr. Stephen Scherer**, Sick Kids Hospital, Toronto, Canada
- **Dr. Nigel Carter** and **Dr. Matthew Hurles**, Sanger Center, UK

➤ Microdeletion/microduplication

- **Dr. Evan Eichler**, Dept. of Genome Sciences, University of Washington, Seattle

➤ Chromosomal array validation

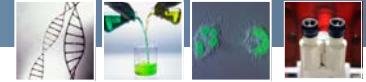
- **Dr. Weimin Bi**, Dept of Human Genetics, Baylor College of Medicine, Houston,

➤ C4 and PSC

- **Dr. Andre Franke**, Institute for Clinical Molecular Biology, Kiel, Germany
- **Dr. Tom Karlsen**, Medical Dept, Rikshospitalet University Hospital, Oslo, Norway

➤ OTC

- **Dr. Allen Bale**, DNA Diagnostics Lab, Yale University



- For Research Use Only. Not for use in diagnostic procedures.
- Practice of the patented 5' Nuclease Process requires a license from Applied Biosystems. The purchase of the TaqMan® Gene Copy Number Assay includes an immunity from suit under patents specified in the product insert to use only the amount purchased for the purchaser's own internal research when used with the separate purchase of an Authorized 5' Nuclease Core Kit. No other patent rights are conveyed expressly, by implication, or by estoppel. For further information contact the Director of Licensing, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA.
- © 2009 Life Technologies Corporation. All rights reserved.
- Trademarks of Life Technologies Corporation and its affiliates include: Applied Biosystems®, AB (Design)®, FAM™ and VIC®.
- TaqMan is a registered trademark of Roche Molecular Systems, Inc.
- All other trademarks are the sole property of their respective owners.