Analysis of copy number variation: design pipeline and validation of TaqMan® Copy Number Assays

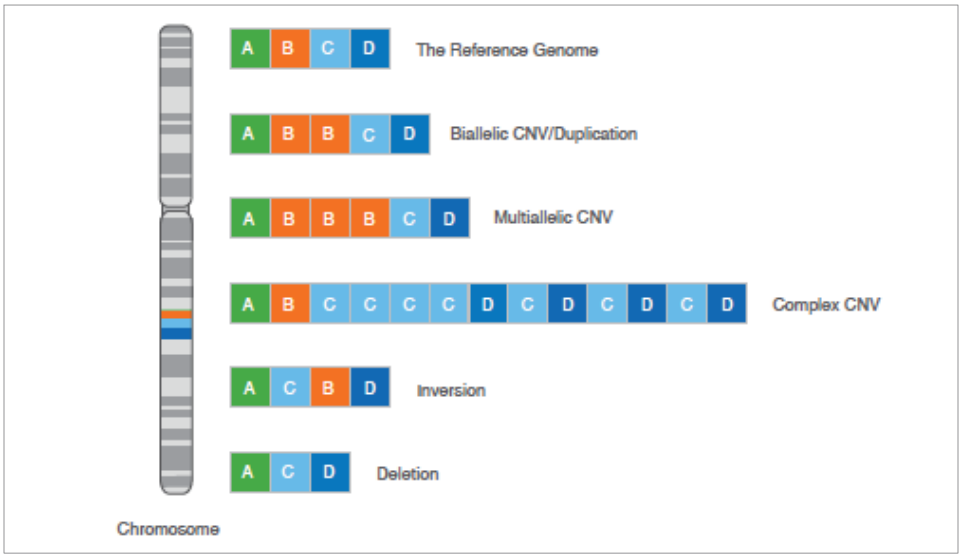

Figure 1. Genomic structural changes affecting segments of DNA result in different types of variations. (Figure adapted from Estivill and Armengol (2007) [1].)

We have designed over 1.8 million TaqMan® Copy Number Assays to address the need for a simple, quantitative method for detection of genomic copy number change, and for validation of copy number change identified using other platforms.

This application note describes the proprietary assay design pipeline that we have developed for TaqMan® Copy Number Assays. The rigorous validation carried out to ensure that the assays produce high-confidence data is also described. The resulting data demonstrate assay accuracy, specificity, and broad application as well as the convenience of using TaqMan® Copy Number Assays for validation of data obtained from copy number discovery platforms. The utility of the assays for screening large sample sizes for specific targets is also shown.

Introduction

Genetic variation in the human genome occurs in many forms—from large, microscopically visible chromosomal aberrations to single nucleotide alterations. Recent studies using microarray-based technologies have revealed the true extent and importance of structural variation within the genome. This variation, known as copy number variation (CNV), covers approximately 12% of the human genome and includes deletions, duplications, and other complex patterns [2], as depicted in Figure 1. These CNVs can influence gene expression

life technologies

and can be associated with specific phenotypes and diseases, as observed in microdeletion and microduplication syndromes [3,4].

Microarray-based technologies are powerful for large-scale CNV discoveries, but they are not the optimum platform for validation of copy number changes and high-throughput screening. A more quantitative technology offering a targeted approach, combined with high accuracy, specificity, ease of use, and sample throughput, is best suited to validate copy number changes or screen large sample sizes.

TaqMan® Copy Number Assays

Overview

We have developed the proprietary TaqMan® Copy Number Assay design pipeline and delivered over 1.8 million pre-designed TaqMan® Copy Number Assays to meet the demand for more quantitative, accurate, and specific copy number detection. These assays target genes (coding and noncoding regions) and intergenic regions, and are useful for screening large numbers of samples and validating data obtained from microarrays and other discovery platforms.

TaqMan® Copy Number Assays enable quantitative detection of DNA copy number using real-time PCR with TaqMan® probe technology, and incorporate a simple workflow. The assays are run in duplex with VIC® dye–labeled TaqMan® Copy Number Reference Assays. These reference genes are known to be present in two copies in a diploid genome, regardless of the copy number of the target of interest, and are used to normalize sample input and minimize the variation between the targets of the test and reference assays.

Assay coverage

Pre-designed TaqMan® Copy Number Assays offer genome-wide coverage encompassing over 95% of RefSeq genes and 85% of copy number variants in the Database of Genomic Variants (DGV; version hg18.v6). There are on average 40 assays per gene and 80 assays per CNV locus, with a median assay spacing of 0.5 to 1 kb.

In the event that a pre-designed assay is not available for a region of interest, Custom TaqMan® Copy Number Assays can be designed. The GeneAssist™ Copy Number Assay Workflow Builder tool incorporates Applied Biosystems® assay design algorithms and facilitates easy online submission of selected genomic regions for custom assay designs.

TaqMan® Copy Number Assay design pipeline

TaqMan® Copy Number Assay probes and primers are designed to target DNA sequences of the genes or genomic regions of interest using a sophisticated, proprietary bioinformatics pipeline, which comprises three main steps (Figure 2).

Step 1—sequence preparation

The public reference genome Build 37 assembly is used as the sequence source for the target regions to which assay probes and primers are designed. The target sequences include genes, known CNV regions, and other regions of the genome. The process involves:

- Masking known single nucleotide polymorphisms (SNPs)
- Masking sequence repeats in the genome
- Identifying the target location
- Ensuring specificity of the target sequences

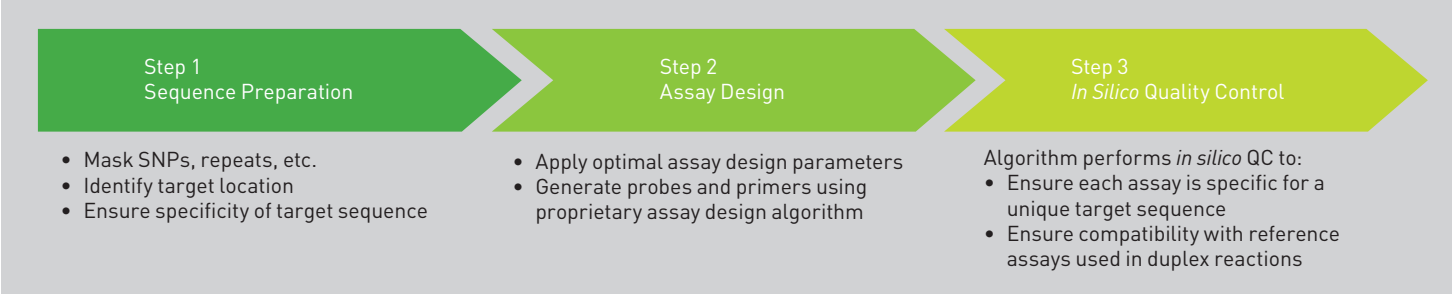

Figure 2. TaqMan® Copy Number Assay design pipeline.

Carrying out these procedures ensures that no probe or primer is designed over ambiguous or known variant nucleotides, and that only unique, high-quality target sequences are put through the pipeline.

Step 2—assay design

Our scientists have empirically determined parameters to select oligonucleotide sequences that are most likely to result in successful amplification. Proprietary software algorithms generate probe and primer designs for the target sequences. To ensure high amplification efficiency, these algorithms include optimal design requirements for parameters such as sequence composition, melting temperature, amplicon length, salt and oligonucleotide concentrations, thermodynamics, and minimal secondary structure.

Step 3—*in silico* QC

This step performs two functions. First, it ensures that each assay is specific to the target sequence for which it was designed (i.e., the assay will not detect sequences from other regions). Secondly, it ensures that each assay is compatible with the TaqMan® Copy Number Reference Assays used in duplex with TaqMan® Copy Number Assays. Assays designed for each target sequence are processed through a quality scoring system, and only the ones that meet the specified criteria are added to the pre-designed TaqMan® Copy Number Assay collection.

Custom TaqMan® Copy Number Assays

For Custom TaqMan® Copy Number Assays, customer-submitted target sequences are run through the same assay design algorithm used to generate the pre-designed assay collection. *In silico* QC, however, does not include a genomic sequence check (i.e., assays are not checked for whether they will detect sequences from other regions). Custom assays can be checked, however, for compatibility with the TaqMan® Copy Number Reference Assays.

Validation of TaqMan® Copy Number Assay design pipeline

To validate the assay design pipeline, over 1,000 assays were randomly selected. TaqMan® Copy Number Assays were tested on a panel of genomic DNA samples, comprising 45 African-American and 45 Caucasian samples, purchased from the Coriell Institute of Medical Research. A subset of these assays was also tested on aneuploid samples, HapMap samples, and on samples with known deletions or duplications [5].

All reactions with TaqMan® Copy Number Assays were performed in duplex using the FAM™ dye label–based assay for the target of interest and the VIC® dye label–based TaqMan® Copy Number Reference Assay for RNase P. Amplification reactions (10 µL), which were performed in quadruplicate, consisted of:

- 10 ng gDNA
- 1X TaqMan® Copy Number Assay
- 1X TaqMan® Copy Number Reference Assay, RNase P
- 1X TaqMan® Genotyping Master Mix

PCR was performed with an Applied Biosystems® 7900HT Fast Real-Time PCR System using the default, universal cycling conditions. Cycle threshold (Ct) values were calculated by the SDS v2.3 software, then relative quantitation was done to estimate copy number for each sample by the ΔΔCt method. Assay performance was evaluated by metrics including call rate, accuracy, and reproducibility. This information was used to set and assess optimal design scoring thresholds for the assay design pipeline. Examples of the tests carried out to validate the TaqMan® Copy Number Assay design pipeline are described here.

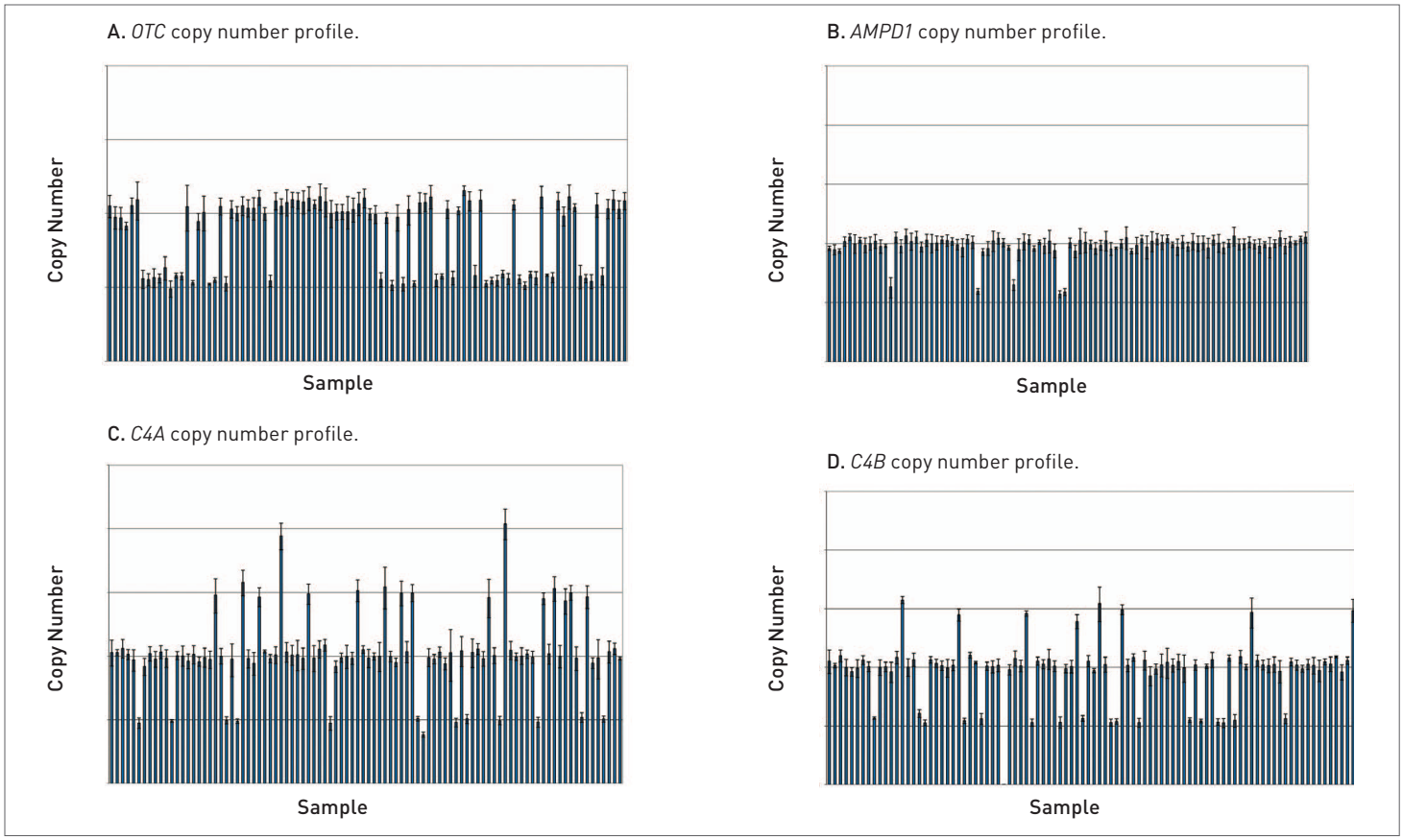

Figure 3. Copy number profiles of selected OMIM genes. TaqMan® Copy Number Assays were used to analyze a panel of genomic DNA samples from African-American and Caucasian populations (45 samples each). (**A**) The copy number profile for the *OTC* gene, located on the X chromosome. All male samples contain one copy, while female samples contain two copies of the *OTC* gene. (**B**) Estimation of *AMPD1* copy number. Deletions were detected in some African-American samples. (**C**, **D**) Distinct copy number profiles for *C4A* and *C4B*, respectively, which encode isoforms of the C4 gene family. Each sample bar represents the mean calculated copy number and error bars show the standard deviation for four replicates.

Copy number determination of OMIM genes

The assays selected to validate TaqMan® Copy Number Assays included some targeted at Online Mendelian Inheritance in Man (OMIM) genes [6]. Four of the copy number profiles for these OMIM genes are shown in Figure 3 as examples of the data obtained during the larger validation study. They are: ornithine carbamoyltransferase (*OTC*), adenosine monophosphate deaminase 1 (*AMPD1*), and isoforms of the complement C4 gene family, *C4A* and *C4B*. A distinct copy number profile was obtained for each gene. For the *OTC* gene, located on the X chromosome, one or two copies were detected in the DNA panel, consistent with the sample gender information (Figure 3A).

For the *AMPD1* gene, no variation in copy number was observed for DNA from individuals in the Caucasian population, but deletions were observed in five individuals from the African-American population (Figure 3B).

The *C4A* and *C4B* genes differ in only five nucleotides within a small 18 bp region of their sequences, making differential detection of copy number very challenging. The proteins encoded by these genes are functionally different, and low copy numbers of the *C4A* gene have been associated with susceptibility to lupus, an autoimmune disorder [7]. As such, it is important that any assay designed to detect either of these genes is able to distinguish between them. TaqMan® Copy Number Assays targeted at these genes were able to discriminate between the two isoforms, displaying distinct copy number profiles for *C4A* and *C4B* (Figure 3C, D).

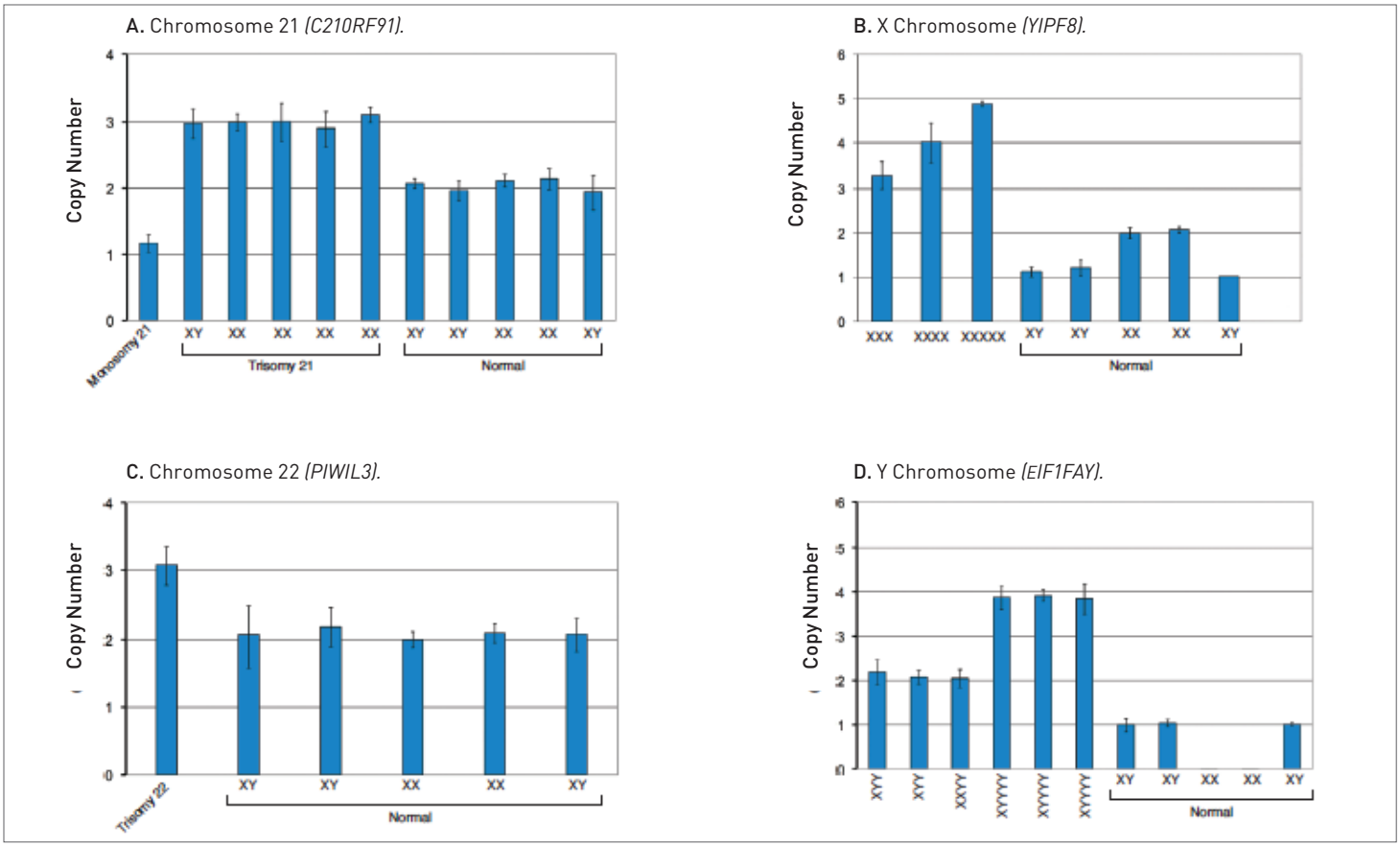

Figure 4. TaqMan® Copy Number Assays accurately identify aneuploidy. TaqMan® Copy Number Assays were designed to target different sequences across each of the chromosomes analyzed. (**A**) Detection of monosomy and trisomy of chromosome 21, with two copies of target detected in normal samples. (**B**) Duplication of the X chromosome is detected in aneuploid samples, while one or two copies of the target are respectively detected in normal samples taken from males and females. For chromosome 22, trisomy is detected in the sample with aneuploidy (**C**) and two copies of the target in samples from normal individuals. (**D**) One copy of the Y chromosome is detected in males, while no Y chromosome target is detected in females. Error bars show the standard deviation of the copy number for the four replicates.

These data demonstrate the excellent performance of TaqMan® Copy Number Assays, including their ability to distinguish between *C4A* and *C4B* genes, which comprise highly homologous sequences.

Accuracy of TaqMan® Copy Number Assays

The ability of TaqMan® Copy Number Assays to accurately estimate copy number was determined using genomic DNA samples of known aneuploidy purchased from the Coriell Institute of Medical Research. Chromosomal aberrations had been previously identified by cytogenetic methods.

Multiple TaqMan® Copy Number Assays were designed across a given chromosome, and used for copy number determination in these aneuploid samples. Normal and aneuploid samples were tested for each assay. The assays were able to accurately determine the copy number of the targeted regions in a given chromosome for each of the samples tested. Examples of one assay for each chromosome are shown (Figure 4). For chromosome 21, one copy was detected in a monosomy 21 sample, while three copies were detected in samples with trisomy 21 (Figure 4A). For the X chromosome, normal samples showed one or two copies of target, depending on gender of samples. Three to five copies of targets were detected in aneuploid samples, consistent with the known chromosomal numbers of the samples (Figure 4B). Three copies of chromosome 22 were detected in a sample with trisomy 22, while two copies were detected in normal samples (Figure 4C). For the Y chromosome, aneuploid samples with two or four Y chromosomes were detected as having two or four copies, respectively. Additionally, one copy of the Y chromosome target was detected in males, and zero copies were detected in females (Figure 4D).

The accuracy of TaqMan® Copy Number Assays was further established in a study using samples of known copy number, performed in collaboration with Dr. E. Eichler at the Department of Genome Sciences, University of Washington. Dr. Eichler and colleagues are studying how structural variation contributes to genomic disorders (e.g., mental retardation, epilepsy, autism, and other congenital anomalies).

To understand the genomic mechanisms behind these disorders, the researchers wanted to map the breakpoints in the DNA of patients with deletions using microarrays. It was necessary to identify individuals with the chromosomal aberration before this could be done. However, these aberrations have an extremely low frequency of approximately 0.2–0.4%, and it would not have been financially feasible or time efficient to screen thousands of samples using microarrays.

To overcome this constraint, five TaqMan® Copy Number Assays were designed to targets across each chromosomal region of interest. Four targets of interest were selected for the study; target regions ranged in size from 500 to 1,200 kb. The regions of interest and their sizes are shown in Table 1. Before the assays could be used to screen the population, however, their accuracy had to be demonstrated. This application note presents the validation data that resulted from a study using samples of known copy number. Positive control samples, depicted as orange bars in Figure 5, had

Table 1. Chromosomal locations of the target sequences used to validate TaqMan® Copy Number Assays.

Aberration	No. of positive samples	Chromosomal coordinates	Size of region (kb)	Assay location
del 17q21.31	4	chr17:41100000-41500000	500	Chr17
dup/del 1q21.1	1	chr1:145100000-145800000	700	Chr1
del 15q24	3	chr15:72300000-73500000	1,200	Chr15_72M
del 15q13.3	1	chr15:29000000-29500000	500	Chr15_29M

previously been identified by other technologies. Four normal samples with two copies of each of the regions analyzed were also included as controls (Figure 5, blue bars). Finally, two copy number assays targeting the X chromosome were used as additional controls. In large sample screens, verifying the gender of a sample can help detect possible sample mix-ups. As well, the results of the X chromosome assays demonstrate that the duplication and deletion events detected are specific to targeted chromosomal regions.

All five TaqMan® Copy Number Assays designed to target each chromosomal region correctly and consistently identified the copy number of the region of interest (Figure 5). All deletions and duplications determined by the TaqMan® Copy Number Assays were consistent with the findings of the previous report that used other technologies to determine the copy numbers of these regions [4]. Two copies were detected in all four normal control samples in each chromosomal region of interest. The copy numbers from the X chromosome control assays were consistent with the gender of each sample. Convinced of the accuracy of the assays, Dr. Eichler and colleagues proceeded to screen over 1,000 patient samples. Four new individuals with the 15q13.3 deletion were identified for further breakpoint mapping analysis, while a further two individuals with the 1q21.1 deletion were also identified.

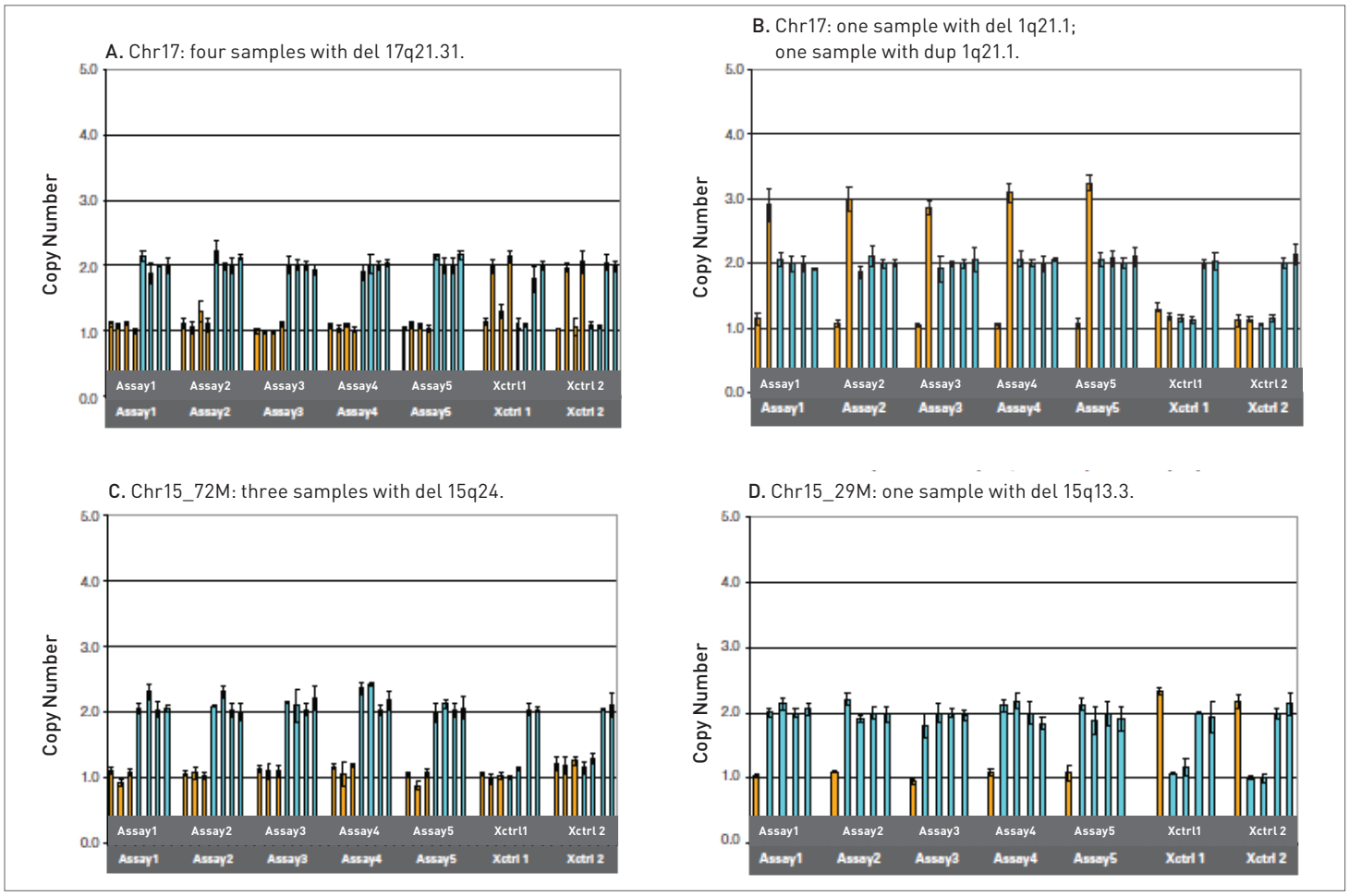

Figure 5. TaqMan® Copy Number Assays are validated by samples with known copy number. Five TaqMan® Copy Number Assays were designed across each chromosomal region of interest. The assays correctly determined the copy number of all positive control samples tested (orange bars). Deletions were observed in all positive samples using assays targeted at assay locations Chr17, Chr15_72M, and Chr15_29M, while assays that targeted Chr1 revealed a duplication in the positive sample. Normal samples known to have two copies of each target region were used as controls (blue bars), as were assays targeting the X chromosome. Positive control samples were previously identified [4]. Error bars show the standard deviation of the copy number for the four replicates.

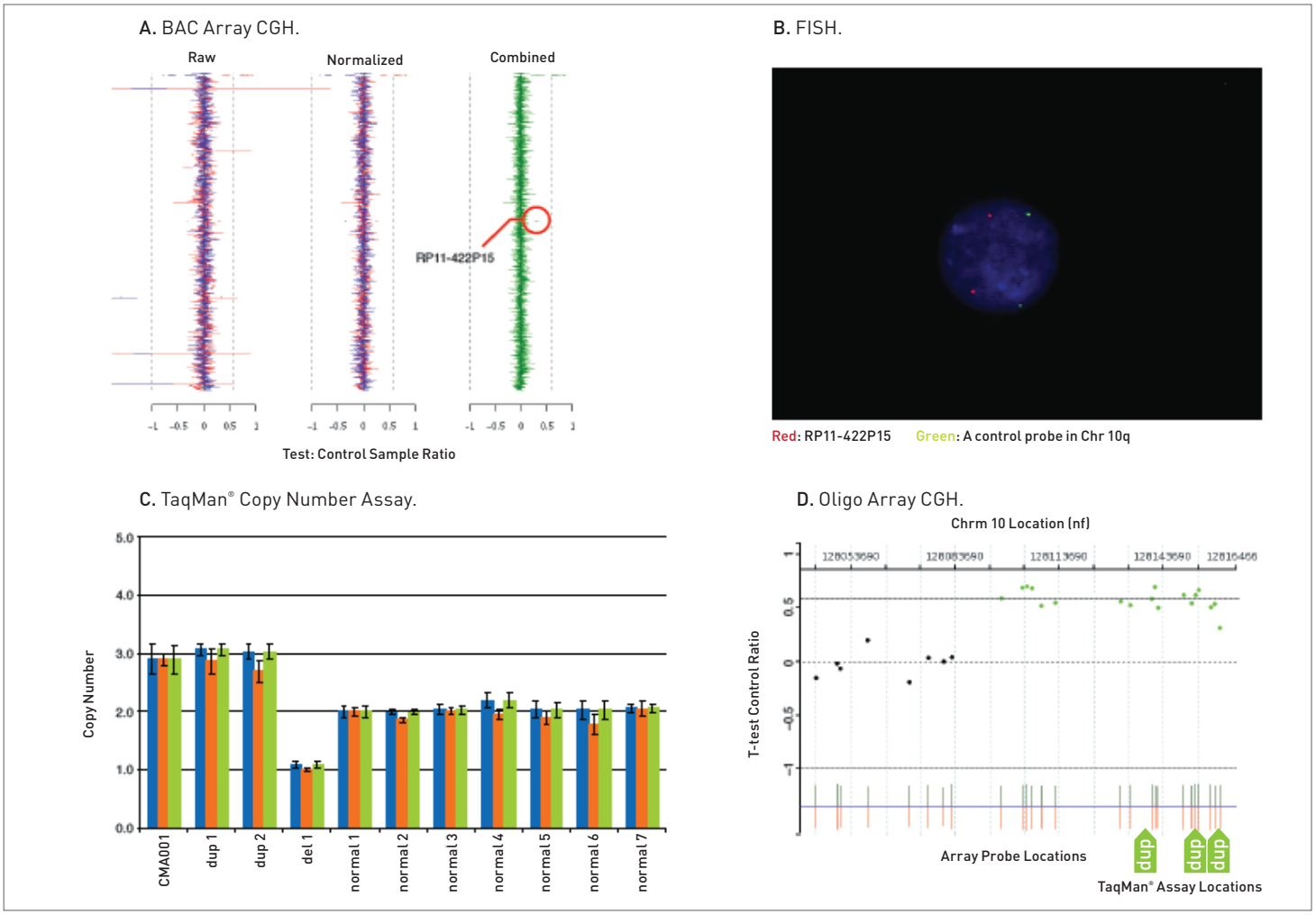

Figure 6. TaqMan® Copy Number Assays clarify ambiguous data from other platforms. (**A**) Identification of duplication in chromosome 10 region 10q26 by BAC aCGH, circled in the green ratio plot (right). (**B**) Subsequent FISH analysis, which did not resolve the duplication. (**C**) Results of three TaqMan® Copy Number Assays targeting different sequences within 10q26. In the copy number bar graph, the three assays are distinguished by color. The CMA001 sample is shown on the left, followed by two control samples containing the duplication, one control sample containing a deletion, and seven normal samples having no CNV in this region. Error bars show the standard deviation of the copy number for the four replicates. (**D**) The duplication in CMA001 was confirmed using oligonucleotide array CGH; the locations of the TaqMan® Copy Number Assays are indicated by arrows.

TaqMan® Copy Number Assays validate data from other technologies

The ability of TaqMan® Copy Number Assays to validate data obtained using other technologies was demonstrated through collaborations with Drs. W. Bi and S. W. Cheung at Baylor College of Medicine, who developed BAC array comparative genomic hybridization (aCGH) methods for prenatal screening of genomic and genetic disorders. Fluorescent *in situ* hybridization (FISH) technologies were used to confirm any positive findings [8].

A gain of copy number was identified in one BAC clone (with sample CMA001), RP11-422P15 at 10q26, using BAC CGH arrays (Figure 6A). However, this duplication was not confirmed by FISH (Figure 6B). Three TaqMan® Copy Number Assays (shown in blue, orange, and green) were designed to target different sequences across this region of RP11-422P15. All three assays were run on a set of samples including CMA001 and various controls. The controls included seven normal samples that contain two copies in the target region. Additionally, two samples containing a duplication in the region and one sample containing a deletion were included. All three TaqMan® Copy Number Assays identified a duplication in the RP11-442P15 region of the CMA001 sample. All copy number calls were consistent across all three assays for each sample (Figure 6C). The duplication in CMA001 was later

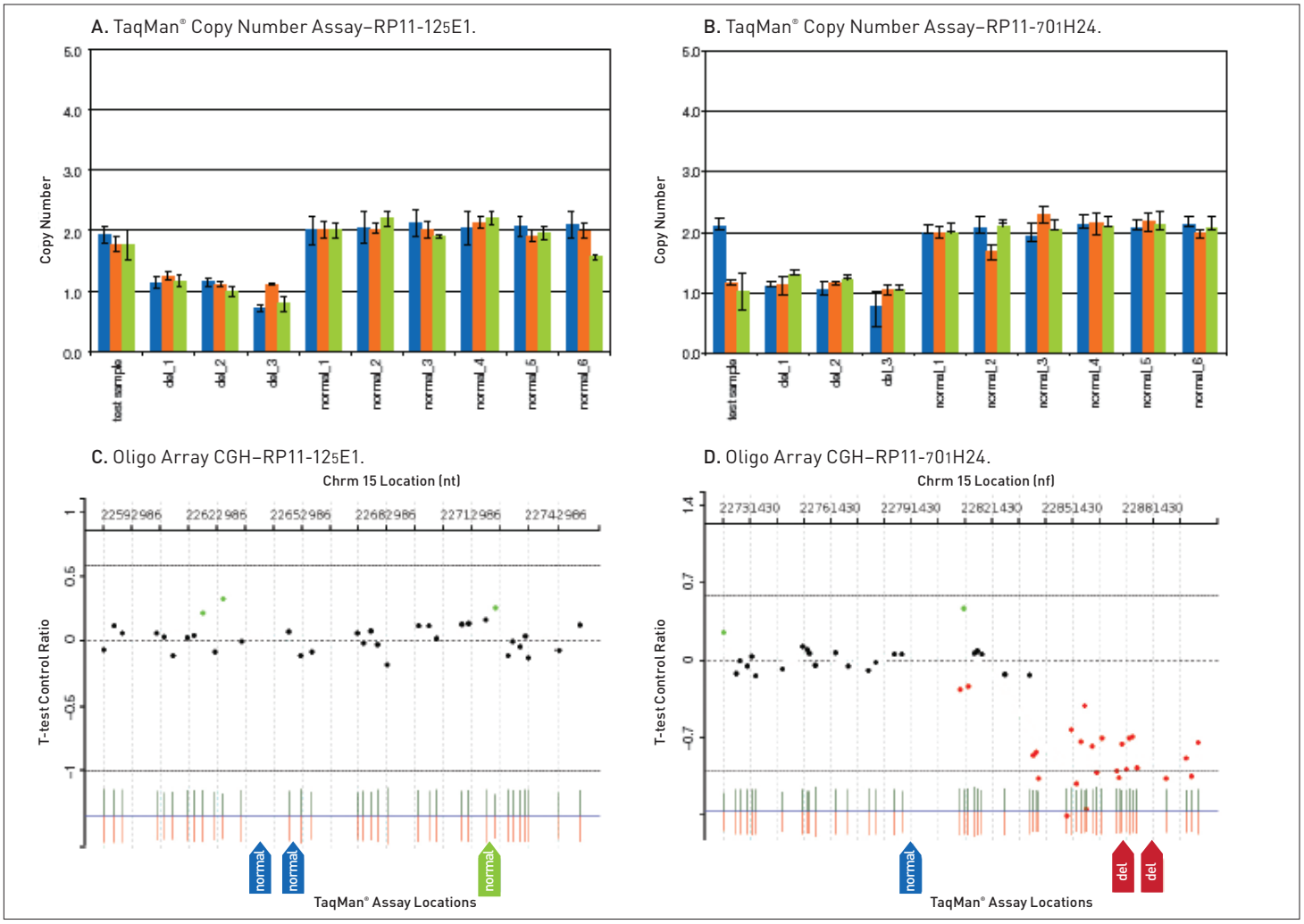

Figure 7. TaqMan® Copy Number Assays validate partial deletions detected by oligonucleotide aCGH. (**A**, **B**) Results of TaqMan® Copy Number Assays targeting three different sequences within chromosome 15 BAC clones RP11-125E1 and RP11-701H24, respectively. The three assays are distinguished by color in the copy number bar graph. In each graph, the CMA003 sample is on the left, followed by three control samples containing deletions and seven normal samples having no CNV in this region. A deletion and deletion breakpoint were detected by the copy number assays in RP11-701H24, but not in RP11-125E1. Error bars show the standard deviation of the copy number for the four replicates. (**C**, **D**) Confirmation of these results using oligonucleotide array CGH; the locations of the TaqMan® Copy Number Assays are indicated by arrows.

confirmed by high-resolution oligonucleotide array CGH (Figure 6D).

For another sample, CMA003, BAC aCGH revealed loss of copy number on chromosome 15, indicated by deletions in BACs RP11-701H24 and RP11-125E1 (data not shown). However, these deletions were not confirmed by FISH. To resolve these ambiguous results, three TaqMan® Copy Number Assays were designed to target regions in both BACs. For each BAC clone region, sample CMA003 was analyzed along with three samples containing a confirmed deletion in the target region and six normal samples with two copies of the targets. No deletion was detected in RP11-125E1 (Figure 7A), but analysis of the downstream adjacent BAC RP11-701H24 indicated that a partial deletion was present and identified the breakpoint within these two clones (Figure 7B). These results were confirmed by an oligonucleotide tiling array; no deletion was detected in RP11-125E1 (Figure 7C), but a partial deletion was detected in RP11-701H24 (Figure 7D).

These results demonstrate that while microarray technologies and FISH may produce conflicting results, TaqMan® Copy Number Assays are able to accurately detect these aberrations, and can clarify ambiguities resulting from using the other technologies.

Simple, accurate copy number variation detection

We have successfully developed a proprietary TaqMan® Copy Number Assay design pipeline that has been rigorously assessed by testing over 1,000 assays for performance. Examples shown in this application note include targets in OMIM genes and in chromosomal regions associated with microdeletion and microduplication syndromes. The data shows that TaqMan® Copy Number Assays can detect copy number changes with high accuracy, as demonstrated by their ability to correctly quantify targets of known copy number in aneuploid samples and other samples with characterized deletions and duplications, and by concordance with results from other platforms. The assays are also highly specific for their target sequences, as demonstrated by their ability to distinguish between highly homologous C4 gene family isoforms that differ by only five nucleotides.

TaqMan® Copy Number Assays incorporate a flexible and simple workflow that can be used for fast analysis of target regions in large sample sets. They also provide a convenient method for validation of CNV discovery results. Furthermore, TaqMan® Copy Number Assays offer a quick means of performing cross-platform comparisons to clarify ambiguous data, taking hours rather than days to complete the experimental analysis.

References

1. Estivill X, Armengol L (2007) Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 3(10):1787–1799.
2. Redon R, Ishikawa S, Fitch KR, et al. (2006) Global variation in copy number in the human genome. *Nature* 444:444–454.
3. Sharp AJ, Mefford HC, Li K, et al. (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* 40:322–328.
4. Mefford H, Sharp AJ, Baker C, et al. (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359:1685–1699.
5. Sharp AJ, Hansen S, Selzer RR, et al. (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 38:1038–1042.
6. Online Mendelian Inheritance in Man. www.ncbi.nlm.nih.gov/omim/
7. Yang Y, Chung EK, Savelli SL, et al. (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80:1037–1054.
8. Cheung SW, Shaw CA, Yu W, et al. (2005) Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med* 7:422–432.

Find out more at **lifetechnologies.com**

life technologies

A Thermo Fisher Scientific Brand