Copy Number Determination Using The SNPlex™ Genotyping System

Michael H Wenz[1], Francisco De La Vega[1], Ryan Koehler[1], Jason Briggs[1], Alexander Spoonde[1], Andreas R Tobler[1], Tom Hemming Karlsen[2], Andre Franke[3], Stefan Schreiber[3], Jochen Hampe[3]

[1]Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA, USA, 94404; [2]Institute of Immunology, Rikshospitalet University Hospital, N-0027 Oslo, Norway; [3]Institute for Clinical Molecular Biology, Christian-Albrechts-University, Schittenhelmstr. 12, 24105 Kiel/Germany.

ABSTRACT

Genome copy number changes (CNPs) are far more frequent than originally expected, and many of them affect gene copy numbers (Aitman et al. Nature 439 (7078), 2006; Sebat et al. Science 305(5683) 2004). Several genetic disorders are the result of CNPs, however because of technical limitations, the extent to which such contribute to phenotypic variations is still poorly understood. We introduced the SNPlex™ Genotyping System to address the need for accurate genotyping data, high sample throughput, study design flexibility, and cost efficiency. The system uses oligonucleotide ligation/polymerase chain reaction (OLA/PCR) and capillary electrophoresis (CE) to analyze single nucleotide polymorphism (SNP) genotypes (Tobler et al. J. Biomol. Tech. 16(4), 2005). Here we demonstrate the feasibility of an adaptation of the SNPlex Genotyping System, to analyze CNPs in a multiplex by comparing the intensity ratios of OLA reactions in test and reference regions. We will present the copy number analysis of DNAs with known chromosomal duplications. Specifically, we studied 88 genomic DNAs, 7 of which contained duplications of the chromosomes 9, 13, 18 or X. On each duplicated chromosome we analyzed at least 10 test OLA reactions, and differences in intensity ratios confirmed all known chromosomal duplications. The assay further identified male (XY) and female (XX) DNA samples due to their copy number difference of the X chromosome.
In addition to identifying known whole-chromosome duplications, we analyzed the copy numbers changes of the RCCX module of the human MHC complement gene cluster as a more complex test case.

INTRODUCTION

Sequence copy number variations in humans range in size from 1kb to 2Mb, are widely distributed throughout the genome and, in several instances, encompass known disease-associated genes. The correlation between gene dosage and disease is well established. Examples include trisomies of chromosomes like 13, 18, and 21 and extensive duplications in the major histocompatibility complex (MHC) region. Many of these variations are polymorphic across populations, termed Copy Number Polymorphisms (CNPs), and it has been suggested that they may contribute to the genetic susceptibility to complex disease.

Assay Principle

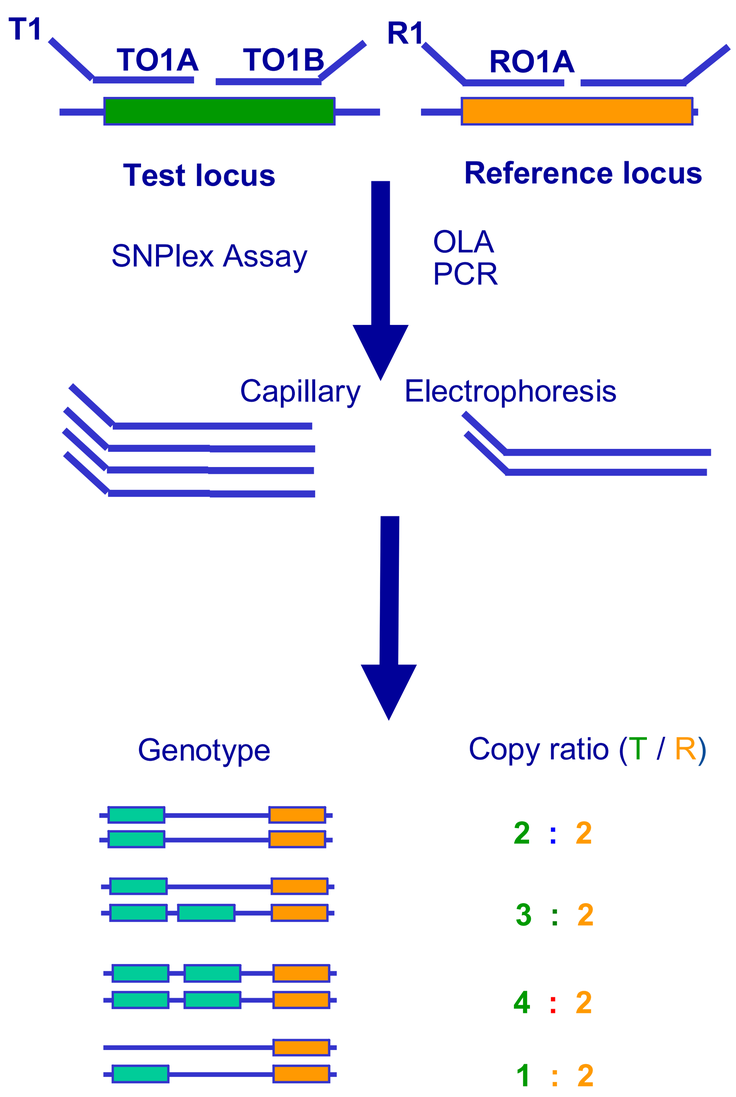

Fig. 1 Assay principle. One OLA reaction targets a reference locus (two alleles), another OLA reaction targets a suspected copy variation sensitive region (test locus). Currently up to 48 reference and test reactions can be performed simultaneously in one multiplex. Quantitative OLA and PCR maintain the ratio between reference and test regions. Peak height information from the data collection software is used to calculate the number of copies of the test sequence, after normalization against the reference sequence.

RESULTS

DETECTION OF TRISOMIC DNA

Chr. 18 Trisomy

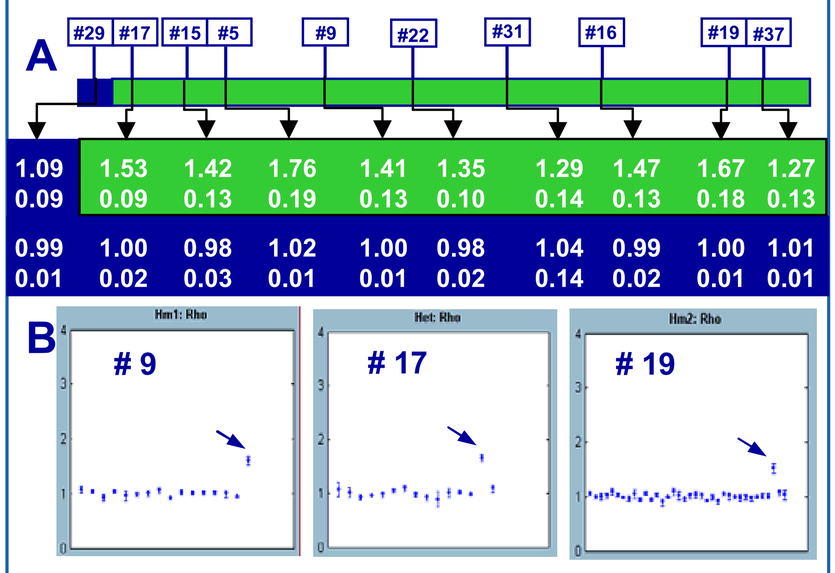

Fig. 2 Copy number analysis of chromosome 18. A: Ten loci (#5-37) across chromosome 18 were addressed by ligation probes. Human non-trisomic DNAs (reference) and one DNA (test; NA01359) were targeted by these probes. The signal ratios and standard deviation obtained after normalization against a reference locus for the test DNA NA01259 (top) and the reference DNAs (bottom) are shown. Green shaded areas indicate loci indicative of three alleles, blue shaded is indicative of 2 alleles representing wild type genotype. In B intensity plots for three of the addresses loci are shown. In each case test DNA NA01359 (arrow) shows an increase in ratio above the reference DNAs.

Chr. X Trisomy

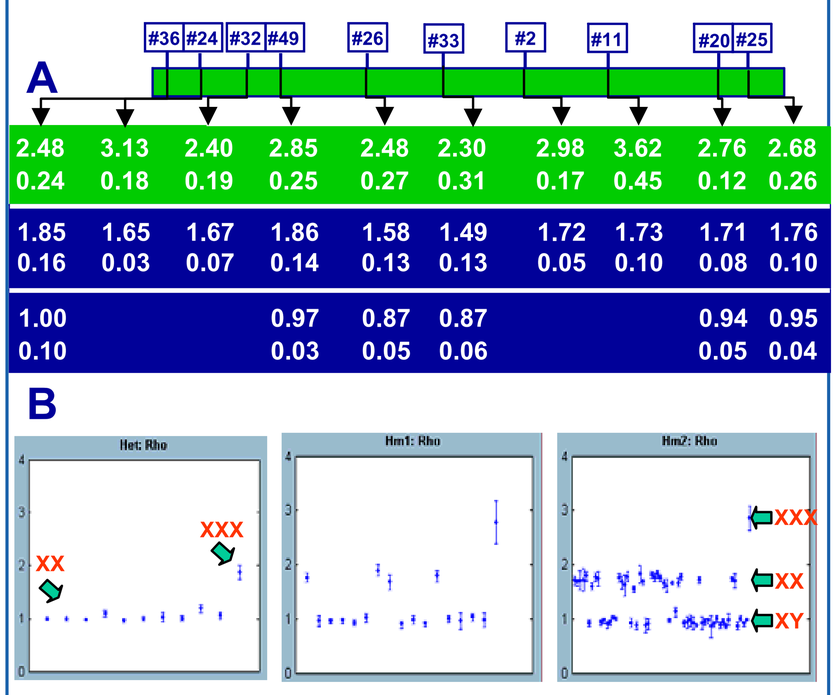

Fig. 3 Copy number analysis of chromosome X. A: Ten loci (#2-49) across chromosome X were addressed by ligation probes. Human male and female reference DNAs and one female test DNA (NA03699) were targeted by these probes. The signal ratios and standard deviation obtained after normalization against a reference locus for the test DNA NA03699) (green, top) and the reference DNAs (blue, bottom) are shown in A.
Representative plots of intensity ratios of test DNA versus reference DNAs are shown in B. Arrows indicate positions for DNAs harboring one copy of chromosome X (male; XY), two copies (female; XX), and three copies (female trisomy, XXX) .

Test DNA	Dupl. Chr.	Test reactions	T/R=
NA03226	9	11	5
NA02948	13	11	10
NA03330	13	11	10
NA01359	18	10	9
NA03623	18	10	10
NA03699	X	11	11
NA04626	X	11	10

Table 1. Test DNA ID, duplicated chromosome, and number of test reactions performed are shown. T/R= indicates number of performed reactions with an increased ratio for the test DNA relative to the reference DNA.

DETECTION OF COPY VARIATION IN THE MHC REGION

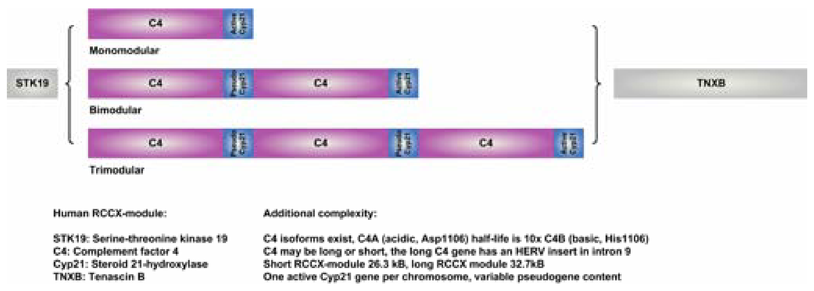

Fig. 4. The complement factor 4 (C4) genes reside within a tightly linked cluster of genes (the RCCX cluster) in the central region of the Human Leukocyte Antigen (HLA) complex on chromosome 6p21. The variation in number and size of C4 genes in the human population is extraordinary, and has been shown to influence protection against infectious disease as well as susceptibility to several autoimmune diseases. Genotyping in the RCCX region for particularly C4 gene content and variation is of great general interest, but has been hampered by the lack of high-throughput methodologies. Incorporating existing knowledge of the region, we have begun to determine the copy number of all functional components of the RCCX module alongside variant copy number for several SNPs and the human endogenous retrovirus indel (HERV-K[C4]) in intron 9.

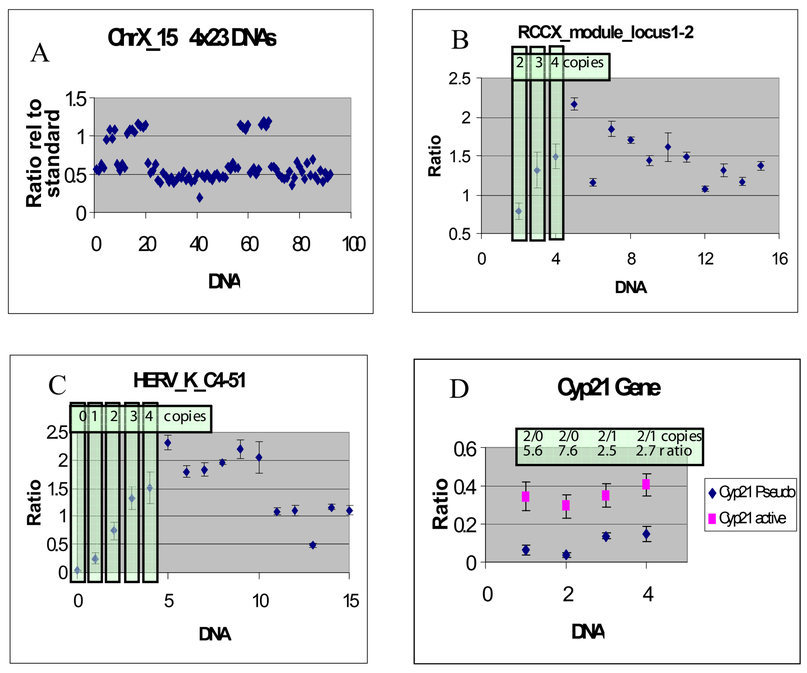

Fig. 5 In a first attempt to address copy number variations in the RCCX cluster, a 48-plex reaction was assembled. Besides addressing multiple targets in the RCCX cluster several probe sets were reserved for addressing control regions outside the HLA complex. The multiplex was tested against a set of 23 genomic DNAs that were plated in quadruplicate on a 96-well plate. For several of those DNAs the copy number for addressed loci were known. (A) shows the identification of the 5 female samples in the DNA set through a probe set targeting a chromosome-X-specific locus. (B) shows copy number ratios for a probe set addressing the RCCX module.
Two of the 23 DNAs were known to harbor 2 copies, 3 DNAs have 3 copies, and 5 DNAs have 4 known copies. Data points are the average for the number of DNAs x replicate platings. (C) shows copy number ratios for a probe set addressing locus HERV-K[C4]. Known copy numbers varied from 0 (2 DNAs), 1 (2 DNAs), 2 (2 DNAs), 3 (4 DNAs), 4 (2 DNAs). (D) shows ratios in gene copy numbers between the Cyp21 active and Cyp21 pseudogene. While the copy number for the active gene is typically 2, the copy number for the pseudogene can vary.

CONCLUSION

We show here initial feasibility of the use of the SNPlex system for the determination of genomic copy number variations (CNVs). The SNPlex system was originally developed for use in high throughput genotyping experiments. Using only slight modifications in the probe design pipeline, the same assay chemistry is utilized. Data analysis is achieved by normalizing for each tested DNA the signal intensity values of a test set against the signal intensity of an unrelated control set. Data presented here indicate that copy number variations between 0 and 3 can be easily discriminated. More complex systems, like the MHC region, where copy numbers can range from 0 to 6, or higher, might require additional tests to confirm the exact number of copies above a threshold of three. The format we present has the advantages of allowing simultaneous typing of both SNPs and CNVs, and the ability to differentiate paralogue duplicated modules.