

Data Mining Empowers the Generation of a Novel Class of Chromosome-specific DNA Probes

H. Zeng^{1,t}, H.U.G. Weier^{1,t}, J. Kwan^{1,5}, M. Wang^{2,6} and B. O'Brien^{1,3,4*}

¹Life Sciences Division, E.O. Lawrence Berkeley National Laboratory, Berkeley, CA, USA

²Human Genome Center, California Institute of Technology (Caltech), Pasadena, CA, USA

³William Harvey Research Institute, Barts and The London School of Medicine, London, UK

⁴Department of Anesthesiology, Berlin Heart Institute, Berlin, Germany

⁵Present address: Albany Medical College, Albany, NY, USA

⁶Present address: Department of Diabetes, Beckman Research Institute of the City of Hope, Duarte, CA, USA

^tThese authors contributed equally to this study

Abstract

Probes that allow accurate delineation of chromosome-specific DNA sequences in interphase or metaphase cell nuclei have become important clinical tools that deliver life-saving information about the gender or chromosomal make-up of a product of conception or the probability of an embryo to implant, as well as the definition of tumor-specific genetic signatures. Often such highly specific DNA probes are proprietary in nature and have been the result of extensive probe selection and optimization procedures. We describe a novel approach that eliminates costly and time consuming probe selection and testing by applying data mining and common bioinformatics tools. Similar to a rational drug design process in which drug-protein interactions are modeled in the computer, the rational probe design described here uses a set of criteria and publicly available bioinformatics software to select the desired probe molecules from libraries comprised of hundreds of thousands of probe molecules. Examples describe the selection of DNA probes for the human X and Y chromosomes, both with unprecedented performance, but in a similar fashion, this approach can be applied to other chromosomes or species.

Keywords: X-linked diseases; Gender determination; Aneuploidy; Cytogenetics; Fluorescence *in situ* hybridization; Prenatal diagnosis; DNA probes; Rational probe design; Data mining; Genomics

Abbreviations: BAC: Bacterial Artificial Chromosome; BLAST: Basic Local Alignment Search Tool; FISH: Fluorescence *in situ* Hybridization; LINE: Long Interspersed Element; NCBI: National Center for Biotechnology Information; OMIM: Online Mendelian Inheritance in Man; UCSC: University of California Santa Cruz

Introduction

Fluorescence *in situ* hybridization (FISH) is an established laboratory method to delineate specific nucleic acid sequences within interphase cell nuclei or metaphase spreads, which has proven indispensable for the detection of specific features in DNA and a variety of chromosomal rearrangements in cytogenetic and cancer research [1-3]. In the typical procedure, non-isotopically labeled nucleic acid probes are prepared first, and hybridized to denatured, single-stranded DNA or RNA targets, before unbound probe molecules are removed by wash steps. The bound probes can be seen in a fluorescence microscope equipped with filter combinations that match the excitation and emission properties of the fluorescent reporter molecules. Several targets can be tagged or 'decorated' simultaneously, and seen as individual hybridization signals, if the probes were labeled with different haptens [4-7].

The studies described below focus on the detection of DNA targets, and more specifically on detection, scoring or enumeration of specific chromosomes in interphase cell nuclei. This type of investigation finds clinical applications in a variety of fields. For example, if one of the parents of an unborn child, embryo or fetus is a known carrier of one of the many known recessive X-linked diseases, the chances of the offspring to be affected by the disease can be better prognosticated by determining its gender [8-10]. This is a relevant issue: according to Online Mendelian Inheritance in Man (OMIM), as of February 22, 2011, the current estimate of sequenced X-linked genes has reached 620 and the total including vaguely defined traits is an estimated 1138 [11].

Another common use of Y chromosome-specific DNA probes is the identification and tracking of grafted cells in sex-mismatched xenograft studies. For example, when Y chromosome bearing white blood cells from a male donor are transplanted to a female recipient, minute amounts of grafted cell can be identified following cell harvest and hybridization due to the presence of the Y-specific signal [12-15].

Materials and Methods

All procedures involving human subject have been reviewed and approved by the UC Berkeley Institutional Review Board.

Retrieval of target sequence for the Y probe

The nucleic acid sequence used for data mining was defined in our previous studies on *in vitro* DNA amplification of Y chromosome-specific DNA repeat sequences. Specifically, we designed pairs of oligonucleotide primers to amplify stretches of 124 bp from the 3.6 kb pentanucleotide DNA repeat described by Nakahori et al. (DYZ1, Genbank accession number X06228) [16,17] (Table 1). Primer annealing sites were chosen that had minimal homology with the human satellite III DNA repeat consensus sequence 'TTCCA' [16,18-

***Corresponding author:** B. O'Brien, M.D., FRCA, William Harvey Research Institute, Translational Medicine & Therapeutics, Charterhouse Square, London, EC1M 6BQ, United Kingdom, Tel: +44 207 882 3402; Fax: +44 207 882 3408; E-mail: benobrien@doctors.org.uk

Received January 21, 2011; Accepted March 14, 2011; Published March 19, 2011

Citation: Zeng H, Weier HUG, Kwan J, Wang M, O'Brien B (2011) Data Mining Empowers the Generation of a Novel Class of Chromosome-specific DNA Probes. J Data Mining in Genom Proteomics 2:108. doi:10.4172/2153-0602.1000108

Copyright: © 2011 Zeng H, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

21]. Blood samples from six normal human volunteers were used to validate the Y chromosome-specific PCR assay [22,23].

Serial cell dilution experiments and artificial mixing of flow-sorted Y chromosome carrying cells in predetermined aliquots of white blood cells from female donors determined that the primer combination WYR 4 (5'-GAACCGTACGATTCCATTTCCTTTTGAA-3') – WYR 6 (5'-TTCCATTCCATTCCATTTCCTTTTCCTTT-3') amplifying a 248 bp DNA fragment corresponding to position 2965-3212 in Genbank accession number X06228 was sufficiently specific to detect a single male cell in the background of 1 million female cells [17,20,23]. Samples comprised entirely of female cells did not yield this product [20-23].

Database searches

We screened the National Center for Biotechnology Information (NCBI) human genome nucleotide DNA database for homologous sequences with one of the most widely used bioinformatics programs, Basic Local Alignment Search Tool (BLAST) [24]. The BLAST approach to rapid sequence comparison directly approximates alignments that optimize a measure of local similarity, the maximal segment pair (MSP) score. The basic algorithm is simple, robust and versatile; it can be implemented in a number of ways and applied in a variety of contexts including straightforward DNA and protein sequence database searches, motif searches, gene identification searches, and in the analysis of multiple regions of similarity in long DNA sequences [24,25].

Execution of the BLAST querying the human genome database with the 27-nucleotide (nt) sequence 'ATTCCGTACGATTCCATTTCCTTTTGAA' from position 3089-3115 of the human Y-specific 3564 bp repeat (Genbank accession number X06228) performed at <http://ncbi.nih.gov/blast> retrieved multiple hits. Parameters were set to identify clones with a range of levels of nucleic

acid homology (setting: 'Optimize for somewhat similar sequences', (BLASTn)).

Retrieval of probe information for the pericentromeric region of the X chromosome

We used the University of California Santa Cruz (UCSC) Genome Browser GRCh37/hg19, built February 2009, at the University of California, Santa Cruz, accessible at <http://genome.ucsc.edu/> to identify bacterial artificial chromosome (BAC) clones with high satellite DNA content [26]. The graphic user interface was set to display BAC end pairs and repeat DNA elements in the pericentromeric region of the human X chromosome, i.e., from position 58,232,531 bp to position 61,922,800 bp.

DNA probe preparation

The BAC DNAs were extracted from overnight cultures following an alkaline lysis protocol [27] or using a ZR BAC DNA Miniprep Kit (Zymo Research; Irvine, CA). The isolation of high molecular weight BAC DNAs was confirmed on 1% agarose gels and quantitated by Hoechst fluorometry using a Hoefer TK 100 instrument (Hoefer; South San Francisco, CA). Probe DNAs were labeled with biotin-14-dCTP (Invitrogen; Carlsbad, CA), digoxigenin-11-dUTP (Roche Molecular Systems; Indianapolis, IN), Spectrum Orange-dUTP (Abbott; Abbott Park, IL) or Spectrum Green-dUTP (Abbott) by random priming using a commercial kit (BioPrime Kit, Invitrogen) [7,28]. When incorporating fluorochrome-labeled deoxynucleoside triphosphates, the dTTP to Spectrum Orange-dUTP or Spectrum Green-dUTP ratio in the reaction was adjusted to 2:1 [28-30].

Fluorescence *in situ* hybridization (FISH) and detection of bound probes

For *in situ* hybridization experiments, 1µl of labeled DNA probe, 1µl

Clone	Band	Start (bp)	End (bp)	Insert size (bp)	Database	Accession numbers
RP11-242E13	Yq12	58819362	58917656	98295	NCBI	AC068123
RP11-294C12	Xq11.1	61696702	61722561	25860	UCSC	AQ508215,AQ508213
RP11-348G24	Xp11.1	58356061	58564667	208607	UCSC	AQ528470,AQ528473

Table 1: BAC clones for human chromosomes X and Y.

Hit	Position in RP11-242E13	Distance to previous hit (bp)
1	3112	Not applicable.
2	6670	3558
3	10234	3564
4	13823	3589
5	17357	3534
6	20831	3474
7	24410	3579
8	27985	3575
9	31579	3594
10	35143	3564
11	38702	3559
12	42276	3574
13	45858	3582
14	49452	3594
15	53026	3574
16	56570	3544
17	60159	3589
18	63713	3554
19	67242	3529
20	70826	3584
21	74400	3574
22	77984	3584
23	81543	3559
24	85107	3564
25	88680	3573
26	92264	3584
27	94258	1994
28	97817	3559

Table 2: Location of 28 BLASTn hits with the 27 nt query sequence in BAC RP11-242E13.

```
gb|AC068123.5|AC068123 Homo sapiens BAC clone RP11-242E13 from Y, complete sequence
Length=98295

Score = 46.1 bits (23), Expect = 0.003
Identities = 23/23 (100%), Gaps = 0/23 (0%)
Strand=Plus/Plus
Query 1      ATTCCGTACGATTCCATTCCTTT 23
            |||
Sbjct 3112    ATTCCGTACGATTCCATTCCTTT 3134

Score = 46.1 bits (23), Expect = 0.003
Identities = 23/23 (100%), Gaps = 0/23 (0%)
Strand=Plus/Plus
Query 1      ATTCCGTACGATTCCATTCCTTT 23
            |||
Sbjct 6670    ATTCCGTACGATTCCATTCCTTT 6692

Score = 46.1 bits (23), Expect = 0.003
Identities = 23/23 (100%), Gaps = 0/23 (0%)
Strand=Plus/Plus
Query 1      ATTCCGTACGATTCCATTCCTTT 23
            |||
Sbjct 10234   ATTCCGTACGATTCCATTCCTTT 10256

Score = 46.1 bits (23), Expect = 0.003
Identities = 23/23 (100%), Gaps = 0/23 (0%)
Strand=Plus/Plus
Query 1      ATTCCGTACGATTCCATTCCTTT 23
            |||
Sbjct 13823   ATTCCGTACGATTCCATTCCTTT 13845

Score = 46.1 bits (23), Expect = 0.003
Identities = 23/23 (100%), Gaps = 0/23 (0%)
Strand=Plus/Plus
Query 1      ATTCCGTACGATTCCATTCCTTT 23
            |||
Sbjct 17357   ATTCCGTACGATTCCATTCCTTT 17379

Score = 46.1 bits (23), Expect = 0.003
Identities = 23/23 (100%), Gaps = 0/23 (0%)
Strand=Plus/Plus
Query 1      ATTCCGTACGATTCCATTCCTTT 23
            |||
Sbjct 20831   ATTCCGTACGATTCCATTCCTTT 20853

Score = 46.1 bits (23), Expect = 0.003
Identities = 23/23 (100%), Gaps = 0/23 (0%)
Strand=Plus/Plus
Query 1      ATTCCGTACGATTCCATTCCTTT 23
            |||
Sbjct 24410   ATTCCGTACGATTCCATTCCTTT 24432
```

Figure 1: The results of a BLASTn query using a human Y chromosome-specific 27nt DNA sequence. Shown are the first seven hits of the alignment search, which in total resulted in 28 perfect matches for 23/27 nt query sequence.

of salmon sperm DNA (10 mg/ml, 5 Prime – 3 Prime Inc.; Boulder, CO), 1µl water, and 7µl of the hybridization master mix (78.6% formamide (FA) (Invitrogen), 14.3% dextran sulfate (Sigma; St. Louis, MO) in 1.43x SSC, pH 7.0) (20x SSC is 3 M sodium chloride and 300 mM trisodium citrate, the pH was adjusted to 7.0) were combined, thoroughly mixed and denatured at 76°C for 10 min. The final concentration in the hybridization mixture is 55% FA, 10% dextran sulfate and 1xSSC, pH 7.0 [29,30]. Acetic acid:methanol (1:3, vol.:vol.) fixed metaphase spreads were prepared on slides from phytohaemagglutinin-stimulated short-term cultures of peripheral blood lymphocytes from a karyotypically normal male following published procedures [31,32]. Just prior to hybridization, slides were denatured for 4 min in 70% FA/2x SSC, pH 7.0 at 76°C, dehydrated in a 70%, 85%, 100% ethanol series, two minutes each step, and air dried. About 2.5µl of hybridization mixture were placed on the denatured samples, covered with a 12 mm diameter circular coverslip, sealed with rubber cement and incubated for 16-20 hours at 37°C in a humidified chamber.

After the rubber cement was carefully removed from the slides, the coverslip was allowed to slip off in 2x SSC at 20°C. Post-hybridization washes were performed as described [7,28] and included two washes in 50% FA/2x SSC buffer at 43°C for 15 min each followed by two washes

in 2x SSC at 20°C [33,34]. If non-fluorescently-labeled DNA probes were hybridized, i.e., the biotinylated or digoxigenin-labeled probes, unspecific binding sites were blocked with PNM buffer (5% nonfat dry milk, 0.1% NaN₃ in PN buffer (0.1M sodium phosphate buffer, pH 8.0, plus 0.05% NP-40)) [7,28] for 10 min, before probes were detected with either fluorescein-conjugated avidin DCS (Vector; Burlingame, CA) or anti-digoxigenin rhodamine (Roche). An additional two washes in 2x SSC were applied to remove unbound avidin or antibodies. Finally, the slides were mounted with 4,6-diamino-2-phenylindole (DAPI) (0.1µg/ml) in antifade solution (0.1% p-phenylenediamine dihydrochloride (Sigma), 0.1x phosphate buffered saline (Invitrogen), 45 mM NaHCO₃, 82% glycerol (Sigma), pH 8.0) and coverslipped.

Image acquisition and analysis

Fluorescence microscopy was performed on a Zeiss Axioskop microscope (Zeiss, New York, NY) equipped with a filter sets for observation of Cy5/Cy5.5, Texas red/rhodamine, FITC or DAPI (84000v2 Quad, ChromaTechnology; Brattleboro, VT). Images were collected using a CCD camera (VHS Vosskuehler; Osnabrueck, Germany) and processed using Adobe Photoshop® software (Adobe Inc.; Mountain View, CA).

Results and Discussion

Chromosome Y BAC clone selection and validation

The bacterial artificial chromosome clone with the highest homology score, RP11-242E13 (Genbank accession number AC068123) contains 28 identical copies of 23 of the 27 nt query sequence in its 98295 bp insert (Figure 1). According to the NCBI database, this clone has been mapped to the long arm of the Y chromosome, band q12 (Table 1). With one exception (hit #27), the 23 bp sites (Figure 1) are spaced in more-or-less regular intervals of slightly more than 3500 bp (Table 2). The site #27 is spaced 1994 bp from the previous site of homology with the query sequence, thus indicating a single incomplete 3.5kb DNA repeat in BAC RP11-242E13. Without including site #27, the sites are spaced an average of 3566 bp, which is in excellent agreement with the published sequence for a single repeat of 3564 bp described by Nakahori et al. [16].

Validation of the Y-specific BAC clone RP11-242E13 by FISH

The result of an *in situ* hybridization experiment using a biotinylated DNA probe prepared from BAC clone RP11-242E13 is shown in Figure 2A. As anticipated, the 98 kb DNA probe decorated the distal long arm of the human Y chromosome suggesting that there are hundreds or even thousands of copies of the complementary DNA target sequence [35] (Figure 2A). Interestingly, even in this simple hybridization experiment that did not include any type of blocking DNA, the probe lit up just the chromosome Y, band q12. No cross-hybridization with other chromosomes that are known to carry satellite III-type heterochromatin such as human chromosome 9 [19] was found. In summary, our findings suggest that BAC clone RP11-242E13 carries multiple copies of the 3.6kb DNA repeat which hybridize exclusively to the human Y chromosome.

BAC clones from the pericentromeric region of the X chromosome

The data mining effort searching for a large insert, recombinant DNA clone that hybridizes specifically to the heterochromatic DNA tandem repeats in the pericentromeric region of the human X chromosome was guided by our prior observation that chromosome-specific alpha satellite DNA repeats can be identified in BAC clones through database searches [36]. We had noticed that alpha satellite DNA repeats which typically appear as chromosome-specific, high order tandem repeats in the pericentromeric region of human autosomes [26,37] can also be found dispersed in low copy numbers in some significant distance from the centromeric regions and *in vitro* DNA amplification can be used to enrich the alpha satellite sequences [36].

The graphical user interface of the UC Santa Cruz Genome Browser suggested that several clones from the RP11 BAC library map into the X chromosome-specific pericentromere and contain alpha satellite DNA repeats (Figure 3). We decided to test two BAC clones, one that is essentially free of interspersed DNA repeats such as short or long interspersed element (LINE) sequences (RP11-294C12) mapping to chromosome X, band q11.2 (Table1) and a second clone (RP11-348G24), which contains a few interspersed DNA repeats and maps to chromosome X, band p11.1 (Figure 3). After retrieval from the freezer, both clones grew in Luria Bertoni broth at approximately the same rate. The DNA yields from 5 ml overnight cultures were similar, but the labeled DNAs showed drastically different hybridization patterns. The results are shown in Figures 2B and 2C.

The digoxigenin-labeled probe prepared from BAC clone RP11-

348G24 containing the interspersed repeats gave a strong signal on the X chromosome, but also unacceptably high levels of cross-hybridization with other autosomes (Figure 2B). Many interphase nuclei exhibited such high levels of cross-hybridization that it was not possible to delineate the X chromosome target. In contrast, hybridization of a biotinylated DNA probe prepared from BAC clone RP11-294C12 resulted in signals that localized exclusively to the pericentromeric region of the human X chromosome (Figure 2C).

To demonstrate that the highly specific DNA probes for chromosomes X and Y can be combined in dual-color multiplex hybridization experiments [38-40], we labeled the Y chromosome-specific DNA from BAC clone RP11-242E13 with Spectrum Green-dUTP (green fluorescence) and the X chromosome-specific DNA from BAC clone RP11-294C12 with Spectrum Orange-dUTP (red fluorescence). Hybridization of this probe mixture gave strong, specific signals in metaphase as well as interphase cells (Figure 2D).

Probe generation and testing rates

Turnaround times for generation and validation of DNA probes prepared from BACs are at least one order of magnitude shorter than those observed with yeast artificial chromosome clones as large insert human genomic DNA probe template [41-43].

Conclusions

This study was undertaken to specifically seek answers to three questions:

1. Can researchers with limited access to computational capabilities data mine and take advantage of the huge resources generated by the International Human Genome Project?
2. How can one identify recombinant, large insert DNA clones with features that make them ideal hybridization probes? and
3. What constitutes an inexpensive and rapid approach to validate the selected clones?

We believe that all three questions have been answered in this study. First, we have been able to demonstrate that simple, publicly available bioinformatics tools such as BLAST searches running on a simple desktop computer allow the operator to data mine and extract the desired information from publicly available archives. Secondly, we began to define rules to predict DNA probe properties based on DNA sequence analysis. Lastly, we were able to show that in a matter of days, chromosome-specific DNA probes can be defined, prepared and validated by FISH.

The importance of probe specificity can not be overemphasized. While recombinant DNA clones carrying large chunks of the human genome such as the BAC or P1 clones [39,40,43-44] are easy to propagate, the presence of non-chromosome-specific interspersed DNA repeats can lead to major impediments in FISH-based interphase cell analysis [36]. As shown above, careful probe selection based on mining of the publicly accessible databases may circumvent some of these problems.

The simple, yet efficient data mining approach presented in this paper is our first step towards creation of an '*in silico*' DNA probe set that may find extensive use in the analysis of NexGen, deep sequencing data [45,46]. Thus, this new class of '*in silico* probes' has little or nothing in common with the conventional recombinant DNA probes [1,2,4,19,47], many of which have not even been sequenced, and it is more akin to the well known '*in silico* PCR' [48]. Here, we describe

a software-based interactive approach in which the user checks the results of the data mining operation and validates the BAC clones via FISH. In our lab setting, the approach allowed us to select and validate a few clones (<10) per week.

In closing, it's important to mention that polymorphisms or centromeric heteromorphisms might present serious problems in the FISH-based analysis of cells from affected individuals using DNA repeat probes [49] regardless whether such an analysis is performed by conventional FISH, Spectral Imaging or multicolor FISH [6,34,50]. The before-mentioned deep sequencing in combination with 'in silico' copy number analysis might prove to be a reasonable alternative approach.

Acknowledgements

This work was supported in parts by a grant from the Director, Office of Energy Research, Office of Health and Environmental Research, U.S. Department of Energy, under contract DE-AC02-05CH11231, the Leonard Rosenman Fund and NIH grants HD45736, CA123370, CA132815, CA132815-02S1 and CA136685 (to HUG).

Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

References

1. Cremer T, Landegent J, Brückner A, Scholl HP, Schardin M, et al. (1986) Detection of chromosome aberrations in the human interphase nucleus by visualization of specific target DNAs with radioactive and non-radioactive in situ hybridization techniques: diagnosis of trisomy 18 with probe L1.84. *Hum Genet* 74: 346-352.
2. Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, et al. (1988) Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci USA* 85: 9138-9142.
3. Lichter P, Ward DC (1990) Is non-isotopic in situ hybridization finally coming of age? *Nature* 345: 93-94.
4. Dauwse JG, Wiegant J, Raap AK, Breuning MH, van Ommen GJ (1992) Multiple colors by fluorescence in situ hybridization using ratio-labelled DNA probes create a molecular karyotype. *Hum Mol Genet* 1: 593-598.
5. Schröck E, du Manoir S, Veldman T, Schoell B, Wienberg J, et al. (1996) Multicolor spectral karyotyping of human chromosomes. *Science* 273: 494-497.
6. Fung J, Weier HU, Goldberg JD, Pedersen RA (2000) Multilocus genetic analysis of single interphase cells by spectral imaging. *Hum Genet* 6: 615-622.
7. O'Brien B, Jossart GH, Ito Y, Greulich-Bode KM, Weier JF, et al. (2010) 'Chromosomal Rainbows' detect oncogenic rearrangements of signaling molecules in thyroid tumors. *The Open Cell Signaling Journal* 2: 13-21.
8. Grifo JA, Tang YX, Munné S, Alikani M, Cohen J, Rosenwaks Z (1994) Healthy deliveries from biopsied human embryos. *Hum Reprod* 9: 912-916.
9. Weier HU, Weier JF, Renom MO, Zheng X, Colls P, et al. (2005) Fluorescence in situ hybridization and spectral imaging analysis of human oocytes and first polar bodies. *J Histochem Cytochem* 53: 269-272.
10. Colls P, Silver L, Olivera G, Weier J, Escudero T, et al. (2009) Preimplantation genetic diagnosis for gender selection in the USA. *Reprod Biomed Online* 19 Suppl 2: 16-22.
11. URL <http://www.ncbi.nlm.nih.gov/Omim/mimstats.html>; OMIM Statistics for February 22, 2011
12. Pallavicini MG, Langlois RG, Reitsma M, Gonzalgo M, Sudar D, et al. (1992) Comparison of strategies to detect and quantitate uniquely marked cells in intra- and inter-species hemopoietic chimeras. *Cytometry* 13: 356-367.
13. Nilsson SK, Hulspar R, Weier HUG, Quesenberry PJ (1996) In situ detection of individual transplanted bone marrow cells using FISH on sections of paraffin embedded murine femurs. *J Histochem Cytochem* 44: 1069-1074.
14. Nilsson SK, Dooner MS, Tiarks CY, Weier HUG, Quesenberry PJ (1997) Potential and distribution of transplanted hematopoietic cells in a nonablated mouse model. *Blood* 89: 4013-4020.
15. Abedi M, Greer DA, Colvin GA, Demers DA, Dooner MS, et al. (2004) Robust conversion of marrow cells to skeletal muscle with formation of marrow-derived muscle cell colonies: a multifactorial process. *Exp Hematol* 32: 426-434.
16. Nakahori Y, Mitani K, Yamada M, Nakagome Y (1986) A human Y-chromosome specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides. *Nucleic Acids Res* 14: 7569-7580.
17. Gray JW, Weier HU (1998) Y chromosome specific nucleic acid probe and method for determining the Y in situ. US Patent number: 5840482, Filing date: Oct 10, 1990, Issue date: Nov 24, 1998.
18. Cooke HJ, McKay RD (1978) Evolution of a human Y chromosome-specific repeated sequence. *Cell* 13: 453-460.
19. Cooke HJ, Hindley J (1979) Cloning of human satellite III DNA: different components are on different chromosomes. *Nucleic Acids Res* 6: 3177-3197.
20. Gray JW, Weier HU (2001) Y chromosome specific nucleic acid probe and method for determining the Y in situ. US Patent number: 6300066, Filing date: Nov 23, 1998, Issue date: Oct 9, 2001.
21. Gray JW, Weier HU (1999) Y chromosome specific nucleic acid probe and method for determining the Y in situ. US Patent number: 5888730, Filing date: Oct 6, 1995, Issue date: Mar 30, 1999.
22. Weier HU, Reitsma M, Gray JW (1990) Detection of fetal cells by *in vitro* DNA amplification. In: *Advances in Analytical Cellular Pathology: Proceedings of the First Conference of the European Society for Analytical Cellular Pathology*, Schloss Elmau, F.R.G., 12-17 November 1989. G. Burger, M. Oberholzer, G.P. Vooijs (eds.), Excerpta Medica, Elsevier Science Publishers B.V., Amsterdam, ISBN:0444813586, Pp. 105-106.
23. Weier HU, Reitsma M, Gray JW (1989) Detection of fetal cells by *in-vitro* DNA amplification. *Analyt Cell Pathol* 1: 313.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
25. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
26. Wayne JS, Willard HF (1985) Chromosome-specific alpha satellite DNA: nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome. *Nucleic Acids Res* 13: 2731-2743.
27. Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513-1523.
28. Weier HU, Rhein AP, Shadravan F, Collins C, Polikoff D (1995) Rapid physical mapping of the human *trk* protooncogene (NTRK1) to human chromosome 1q21-q22 by P1 clone selection, fluorescence in situ hybridization (FISH), and computer-assisted microscopy. *Genomics* 26: 390-393.
29. Weier HUG, Wang M, Mullikin JC, Zhu Y, Cheng JF, et al. (1995) Quantitative DNA fiber mapping. *Hum Mol Genetics* 4: 1903-1910.
30. Fung J, Hyun W, Dandekar P, Pedersen RA, Weier HUG (1998) Spectral imaging in pre-conception/preimplantation genetic diagnosis (PGD) of aneuploidy: multi-colour, multi-chromosome screening of single cells. *J Ass Reprod Genet* 15: 323-330.
31. Fung J, Weier H-UG, Pedersen RA, Zitzelsberger HF (2002) Spectral imaging analysis of metaphase and interphase cells. In: *FISH Technology*. B. Rautenstrauss and T. Liehr (Eds.), Springer Verlag, Heidelberg, Pp.363-387.
32. Bayani J, Squire JA (2004) Preparation of cytogenetic specimens from tissue samples. In: Bonifacino JS, ed. *Current Protocols in Cell Biology*, Suppl 23. New York, John Wiley & Sons, Inc. Pp. 22.2.1-22.2.15.

33. Kwan J, Baumgartner A, Lu C-M, Wang M, Weier JF, et al. (2009) BAC-FISH assays delineate complex chromosomal rearrangements in a case of post-Chernobyl childhood thyroid cancer. *Folia Histochem Cytobiol* 47: 135-142.
34. Lu C-M, Kwan J, Baumgartner A, Weier JF, Wang M, et al. (2009) DNA probe pooling for rapid delineation of chromosomal breakpoints. *J Histochem Cytochem* 57: 587-597.
35. Weier HU, Seagraves R, Pinkel D, Gray JW (1990) Synthesis of Y-chromosome specific, labeled DNA probes by *in vitro* DNA amplification. *J Histochem Cytochem* 38: 421-427.
36. Baumgartner A, Weier JF, Weier HUG (2006) Chromosome-specific DNA repeat probes. *J Histochem Cytochem* 54: 1363-1370.
37. Wayne JS, Willard HF (1989) Concerted evolution of alpha satellite DNA: evidence for species specificity and a general lack of sequence conservation among alphoid sequences of higher primates. *Chromosoma* 98: 273-279.
38. Jossart GH, O'Brien B, Cheng JF, Tong Q, Jhiang SM, et al. (1996) A novel multicolor hybridization scheme applied to localization of a transcribed sequence (D10S170/H4) and deletion mapping in the thyroid cancer cell line TPC-1. *Cytogenet Cell Genet* 75: 254-257.
39. Bednarz N, Eltze E, Semjonow A, Rink M, Andreas A, et al. (2010) BRCA1 loss preexisting in small subpopulations of prostate cancer is associated with advanced disease and metastatic spread to lymph nodes and peripheral blood. *Clin Cancer Res* 16: 3340-3348.
40. Fung J, Weier H-UG, Pedersen RA (2001) Detection of structural and numerical chromosome abnormalities in interphase cells using Spectral Imaging. *J Histochem Cytochem* 49: 797-798.
41. Fung J, Munné S, Duell T, Weier HUG (1998) Rapid cloning of translocation breakpoints: from blood to YAC in 50 days. *J Biochem Mol Biol Biophys* 1: 181-192.
42. Lu C-M, Kwan J, Weier JF, Baumgartner A, Wang M, et al. (2009) Rapid mapping of chromosomal breakpoints: from blood to BAC in 20 days. *Folia Histochem Cytobiol* 47: 367-375.
43. Fung J, Munné S, Garcia J, Kim UJ, Weier HUG (1999) Reciprocal translocations and infertility: molecular cloning of breakpoints in a case of constitutional translocation t(11;22)(q23;q11) and preparation of probes for preimplantation genetic diagnosis (PGD). *Reprod Fertil Devel* 11: 17-23.
44. Shepherd NS, Pfrogner BD, Coulby JN, Ackerman SL, Vaidyanathan G, et al. (1994) Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc Natl Acad Sci USA* 91: 2629-2633.
45. Kim YJ, Teletia N, Ruotti V, Maher CA, Chinnaiyan AM, et al. (2009) ProbeMatch: rapid alignment of oligonucleotides to genome allowing both gaps and mismatches. *Bioinformatics* 25: 1424-1425.
46. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760.
47. Greulich-Bode KM, Wang M, Rhein AP, Weier JF, Weier HU (2008) Validation of DNA probes for molecular cytogenetics by mapping onto immobilized circular DNA. *Mol Cytogenet* 1: 28.
48. Ho SM, Tang WY (2007) Techniques used in studies of epigenome dysregulation due to aberrant DNA methylation: an emphasis on fetal-based adult diseases. *Reprod Toxicol* 23: 267-282.
49. Weier HUG, Gray JW (1992) A degenerate alpha-satellite probe, detecting a centromeric deletion on chromosome 21 in an apparently normal human male, shows limitations of the use of satellite DNA probes for interphase ploidy analysis. *Anal Cell Path* 4: 81-86.
50. Greulich KM, Kreja L, Heinze B, Rhein AP, Weier HG, et al. (2000) Rapid detection of radiation-induced chromosomal aberrations in lymphocytes and hematopoietic progenitor cells by FISH. *Mut Res* 452: 73-81.