A Contiguous 3-Mb Sequence-Ready Map in the S3–MX Region on 21q22.2 Based on High-Throughput Nonisotopic Library Screenings

Thomas Hildmann,1 Xianging Kong,1 John O’Brien, Lisa Riesselman, Hoang-My Christensen, Emilie Dagand, Hans Lehrach, and Marie-Laure Yaspo2

Max Planck-Institut für Molekulare Genetik, D-14195 Berlin-Dahlem, Germany

Progress in complete genomic sequencing of human chromosome 21 relies on the construction of high-quality bacterial clone maps spanning large chromosomal regions. To achieve this goal, we have applied a strategy based on nonradioactive hybridizations to contig building. A contiguous sequence-ready map was constructed in the Down syndrome congenital heart disease (DS-CHD) region in 21q22.2, as a framework for large-scale genomic sequencing and positional candidate gene approach. Contig assembly was performed essentially by high throughput nonisotopic screenings of genomic libraries, prior to clone validation by (1) restriction digest fingerprinting, (2) STS analysis, (3) Southern hybridizations, and (4) FISH analysis. The contig contains a total of 50 STSs, of which 13 were newly isolated. A minimum tiling path (MTP) was subsequently defined that consists of 20 PACs, 2 BACs, and 5 cosmids covering 3 Mb between D21S3 and MX1. Gene distribution in the region includes 9 known genes (c21–LRP, WRB, SH3BGR, HMG14, PCP4, DSCAM, MX2, MX1, and TMPRSS2) and 14 new additional gene signatures consisting of cDNA selection products and ESTs. Forthcoming genomic sequence information will unravel the structural organization of potential candidate genes involved in specific features of Down syndrome pathogenesis.

Research on Down syndrome (DS) has been the driving force behind molecular genetics analysis of human chromosome 21. A tremendous mapping effort was put forward, and YAC maps spanning most of 21q were established (Chumakov et al. 1992; Nizetic et al. 1994; Gardiner et al. 1995). Down syndrome is caused by trisomy 21, and the characteristic phenotype reflects the consequences of dosage imbalance for a set of genes localized on chromosome 21 (Epstein et al. 1991). The disease affects 1 in 700 live births and is the major cause of mental retardation and congenital heart disease. Rare documented cases of partial trisomy 21 allowed the mapping of specific DS phenotypic features into discrete chromosomal segments (Delabar et al. 1993; Korenberg et al. 1994). One of these candidate regions, the so-called DCR (DS critical region) that is associated with characteristic mental retardation and skeletal dysmorphies (Rahmani et al. 1989), has been a popular focus of investigation (Dahmane et al. 1995; Ohira et al. 1996, 1997; Osoegawa et al. 1996). However, the DCR boundaries are difficult to assess, owing in part to the variable and incomplete penetrance of most DS traits (Hernandez and Fisher 1996). Furthermore, there is significant evidence that other regions contribute to the complex phenotype (Korenberg et al. 1994). Congenital heart defects of the endocardial cushion (DS-CHD) and congenital gut conditions have been associated with a 5-Mb region in 21q22.2 encompassing D21S55 to MX (Korenberg et al. 1992). DS-CHD affects >40% of the DS patients. Genotype–phenotype correlation has prompted further mapping initiatives (Hubert et al. 1997) as a step toward candidate gene identification in the DS-CHD region.

Selection of positional candidates in large chromosomal regions requires extensive isolation of gene signatures. Gene discovery on chromosome 21 has taken a new turn in 1997 with the launching of a genomic sequencing program shared by a consortium of laboratories in Japan and Germany (Gardiner and Yaspo 1998). The whole long arm of chromosome 21 is scheduled to be sequenced by year 2000. Within this framework, construction of physical maps that achieve contiguity yet with minimum clone overlaps is a critical issue. To build long-range contigs adequate for genomic sequencing, we have applied a strategy based on high throughput nonisotopic screening of arrayed genomic libraries. We describe here a gapless sequence-ready map spanning 3 Mb between D21S3 and TMPRSS2 that is currently in the sequencing phase.

RESULTS

Strategy Used for Contig Assembly

Large-scale nonisotopic screenings of genomic libraries were used as a means to build contigs by hybridization. The strategy that we have applied was the following:
(1) assembling primary contigs by high throughput hybridization of high-density grids; (2) gap-filling by multipoint walkings; and (3) clone fingerprinting by restriction digest.

**Primary Contig Assembly**

In a first step, a raw primary contig was assembled by multipoint high throughput screening of arrayed PAC (Ioannou et al. 1994) and cosmid libraries (Nizetic et al. 1991) using available markers and anchors previously localized in the targeted region (see Methods). We initially used 24 known STS markers for screening the genomic libraries. Although STSs provided one marker per 125 kb (24 STSs for 3 Mb) on average, the marker distribution across the region was uneven. In particular, the sole STS localized between D21S345 and D21S15, namely D21S347, was in fact mapping to 21q21 (data not shown), leaving a 700-kb gap between the flanking STSs. In an attempt to compensate for the STS distribution bias, we generated novel anchor points from ends of cosmids previously reported in this region: 25 cosmids from the YAC pocket map (pockets 125–132) (Nizetic et al. 1994) and 16 cosmids binned to YACs 767D6 and 784H7 (Soeda et al. 1995). All together, initial library screenings performed with 65 anchor probes derived from STSs and cosmid riboprobes identified a total of 1023 PACs and 956 cosmids. Filters were scored semiautomatically using the Xdigitize program keeping record of the relative signal intensity, referred to as either high, medium, or faint. A typical riboprobe hybridization on a PAC high-density grid is illustrated in Figure 1, showing in this case two positives. After the scoring of multiple hybridization experiments, data files were processed with the well2clone program. The well2clone package was designed for managing large hybridization data sets, pro-

![Figure 1](image.png)
Figure 2. Contig and gene map between D21S3 and TMPRSS2. The sequenced MTP consists of two BACs, 20 PACs, and five cosmids depicted in red. Sizes of clones are indicated in parentheses. Markers are shown above the clone map: STSs are displayed in the lower layer; EST markers are in blue. NotI linking clones are marked with a dot. Riboprobes are shown in the upper layer; riboprobes in magenta correspond to the initial cosmid anchors (pockets are prefixed by ICRF and Soeda’s bins by Q). Riboprobes used for Southern hybridizations to confirm the contig are shown as follows: DIG riboprobes are marked by a green dot and 32P-labeled riboprobes by a red square at the end of corresponding clones. Known genes are depicted by blue lines below the clones. Transcription units (TUs) are indicated below the PACs in magenta. The raw primary data are shown below the actual MTP and gene map. Positive clone names are given for probes showing <20 hits. Color boxes correspond to probes identifying >20 clones for which only numbers are given (cosmids are indicated by a C and PACs by a P). Box color shadings are as follows: 20–50 hits are light yellow, 50–100 hits are yellow, 100–200 hits are dark yellow, and >200 hits are brown. Each probe is indicated above the clone stack.
providing also a dedicated database (Mott et al. 1993). These algorithms were applied previously in several mapping initiatives using high-density grids (Hoheisel et al. 1993; Nizetic et al. 1994). The initial function of well2clone is to merge the results of all single hybridization experiments held in separate files (each containing data of a given probe) into a large flat file. Then, the “probeorder” function allowed the ordering of probe and positive clones from the whole data set using simulated annealing, with the assumption that all probes are single copy. Finally, given an order of probes initially imposed by the user according to known STSs, the “reorder” function gave the best fitting order of clones. As a result, a contig file was calculated, containing all the positive hits that were, in practice, arranged in a series of stacks below their corresponding probes. Connections between clones were showed by “annealed clone groups,” hence suggesting possible contigs. Hybridization along the whole S3–MX region indicated that the number of hits per probe was clearly locus and probe dependent (Fig. 2), which may reflect the uneven representation of a given locus in the genomic libraries. For constructing the whole map (see below), we used a total of 112 probes. Out of these, 21 probes (19%) identified >50 clones in the PAC library (Table 1), possibly owing to low-copy repeats or paralogous sequences. This effect was particularly pronounced in the PAC library that is not chromosome specific. We identified a set of nonisotopic screenings (88 riboprobes and 24 STSs) were used for constructing the whole map. Data are summarized in Table 1. Coverage with cosmid PACs was confirmed to be real positives, reflecting the relative specificity of the initial probe. Clone overlaps were primarily determined from shared STSs, from which we deduced intervals between markers. At this stage, further rounds of hybridizations were performed with riboprobes derived from PCR-validated clones to (1) identify overlaps, (2) orient clones, and (3) fill the gaps. The 700-kb gap between D21S15 and D21S345 was filled by a walking strategy using riboprobes located at the outer edge of existing contigs, 70124-SP6 and 171F15-T7 (Fig. 2). PAC 267O10, identified by 171F15-T7, was used as a probe and did hit PAC 58D10. A further walk with 58D10-SP6 identified PAC 31P10. Walking steps from the centromeric side of the gap (70124-SP6) identified a path of clones confirmed by Southern blot hybridization (not shown) where finally the cosmid Q7G4 and Q82G10 did bridge BAC 291B3 and PAC 31P10 (Fig. 2). Another gap of ~100 kb was detected between HMG14 and D21S346. Two riboprobes derived from clones flanking this gap (128M19-SP6 and 191P10-SP6) simultaneously identified a single PAC (103P10-T7); the bridge was confirmed by PCR and Southern blotting (not shown).

In summary, a total of 112 nonisotopic screenings (88 riboprobes and 24 STSs) were used for constructing the whole map. Data are summarized in Figure 2, showing the clone coverage at each locus. PAC coverage along the whole region was 21 times on average (431 clones with a mean size of 150 kb), excluding from this figure the 21 probes that gave noisy hybridization (shown in Table 1). Coverage with cosmid clones was 11 times.

Thirteen new STSs were generated from end sequences derived from one BAC, five PACs, and two cosmids (Table 2). Additional markers were also tested on the contig: four ESTs (SGC31654, WI-18279, WI-9587, and WI-11358), PCP4 (Chen et al. 1996), and seven recently published STSs (28F9-SP6, 31P10-SP6, 31P10-T7, 737E2-SP6, 628H2-T7, 539E7-SP6, and 539E7-T7) (Hubert et al. 1997). Table 3 summarizes all STS screening data corresponding to the map shown in Figure 2.

**FISH**

Sixty-three clones were analyzed by FISH on metaphase chromosomes during the course of this study (data not shown); 55 clones (87%) mapped to chromosome 21 at a single locus. Six clones mapped to different chromosomes and two clones were localized to a proximal location on chromosome 21.

**Restriction Fingerprinting Analysis**

Contig construction using the well2clone software has...
the inherent limitation that sizes of clones and overlaps cannot be displayed. Fingerprinting was thus performed for the 55 fished clones, using EcoRI or HindIII, as illustrated in Figure 3A. Gel images were digitized and processed with the Image 3 and FPC programs (Sulston et al. 1989; Soderlund et al. 1997) to estimate the clone sizes and overlaps by analysis of the fingerprinting patterns. We observed that clone sizes were generally underestimated by 8% in EcoRI gels and by 18% in HindIII gels, as compared with the size determined by genomic sequencing (data not shown). This was owing in part to the fact that restriction fragments <500 bp were not included in the FPC analysis. Also, fragments comigrating with other insert bands or with vector fragments (two fragments for EcoRII and four for HindIII) may be omitted. Different vector-insert junctions obtained with these enzymes also contribute to the size discrepancy seen between EcoRI and HindIII patterns. Clone sizing by pulsed-field gel electrophoresis (PFGE) after NotI digestion was the most accurate, with a variation of only 4% from sequencing data (not shown). The estimation of clone overlaps calculated by FPC is presented in Table 4. Discrepancies observed in clone sizes are reflected in the estimation of overlaps. For the 12 pairs of clones that were digested with both enzymes, the overlap may vary from 10% to 50% between EcoRI and HindIII estimations. For the 26 pairs of clones representing the final sequenced MTP (marked with “x” in Table 4), only three overlaps were found by FPC, whose sizes were consistent with genomic sequence data. For the other clones, evidence of overlaps was confirmed independently by Southern hybridization or PCR. Data indicate that when using only EcoRI or HindIII and under the parameters used (cutoff probability of 10^{-17}), FPC failed to detect clone overlap <30 kb. The large 65-kb overlap for pair 31P10–58D10 was detected by FPC but above the threshold of 10^{-17} (Table 4). Overlaps determined by genomic sequencing ranged from 4 kb to 80 kb but was generally below 20 kb.

Only four of the PACs analyzed by restriction digest were found grossly deleted, as detected by abnormal restriction pattern. STS content and riboprobe hybridization data indicated that PAC 191P10 had suffered an internal deletion.

### Transcript Map

Localization and orientation of nine known genes reported previously in this region are shown in Figure 2. We have also localized nine additional potential transcripts grouped in six transcriptional units (TU84,
TU85, TU117, TU122, TU149, and TU188) that were isolated previously in the course of constructing a whole chromosome 21 transcript map (Yaspo et al. 1995; M.L. Yaspo, unpubl.). Essentially, the transcript map consists of gene fragments originating from cDNA selection (Parimoo et al. 1991; Korn et al. 1992) and exon trapping (Church et al. 1994) that have been systematically mapped to human genomic libraries (Yaspo et al. 1995). TU188 (EMBL accession no. AJ003597) shows 60% homology and 42% identity with β-1,3-galactosyltransferases. cDNA MPIpl12-2I8 (EMBL accession no. AJ003472, 355 bp) shows 50% homology over its whole length with Titin genes and is likely to be part of DSCAM, also displaying a regional homology with Titin (Yamakawa et al. 1998). We have also re-mapped and ordered five STSs from the Unigene EST set (Schuler et al. 1996). It is worth mentioning that no additional transcripts were mapped in the 700-kb region spanned by DSCAM (Fig. 2).

**DISCUSSION**

A nonisotopic hybridization-based strategy was applied for building a contiguous contig spanning 3 Mb from D21S3 to MX. Semiautomated hybridization and detection procedures enabled the managing of high throughput screenings and data processing. The well2clone program was compulsory for the efficient
managing of this large data set and for establishing
table 2 in the later stages of contig construction. The clone depth
was >20 times over the whole region with a dense distri-
bution of probes, thus giving a prospect for choosing
suitable clones.

Nonisotopic screenings represent a substantial ad-

vance in the setup of high throughput mapping. The
contig built by hybridization was then confirmed by
applying independent analysis methods to define a
minimum tiling path (MTP) for genomic sequencing.
Clones were anchored to their neighbors on the basis
of shared STSs, fingerprinting, and Southern blot hy-
bridizations. Thirteen novel STSs have been generated,
and the contig contains a total of 50 STSs including
NotI linking clones, hence anchoring the contig to the
NotI physical map (Ichikawa et al. 1992). This provides
a valuable backbone allowing the estimation of real
distances between markers. However, hybridization-
based contig construction does not give the size of
clones and overlaps, which were estimated here by fin-
gerprinting. Using FPC as a means to calculate whether
a pattern of fragments common to two clones is sig-
ificant or not, we set the probability cutoff to <10^{−7}.

Discrepancies between values obtained for EcoRI and
HindIII patterns were seen in 6 of 12 pairs of clones for
which data could be compared. Besides, only 3 over-
laps were found significant among the 26 pairs of over-
lapping clones chosen for the MTP. When compared
with sequencing data available for 23 of 26 pairs, over-
laps ranged from 2 to 32 kb, except for 5 pairs showing
larger values. Actual overlaps range between 1% and
30% of the clone length, with a mean clone to clone
overlap of 13% over the whole map. As a conclusion,
analyzing patterns from only one enzyme, either EcoRI
or HindIII, was not an adequate choice for detecting
overlaps <30 kb. Also, the parameters set for FPC may
have been too stringent; when considering a proba-
bility cutoff <10^{−5}, nine additional overlaps are detected
(Table 4). A more complex restriction pattern would be
an advantage, for instance, using these two enzymes in
combination. In contrast to other studies (Mara et al.
1997), the contribution of FPC in our hybridization-
based approach as not used as a means to define ten-
tative contigs, thus limiting the impact of nondetected
overlaps. The overall strategy was proven adequate to
construct a sequence-ready map consisting of 20 PACs,
2 BACs, and 5 cosmids. The use of two cloning systems,
PAC and cosmids, gave a chance to identify three “cos-
smid bridges” between PACs, hence minimizing se-
quencing redundancy. The average clone to clone
overlap is 13% over the whole region.
As compared with an STS-based BAC contig published previously and encompassing the same region (Hubert et al. 1997), we have closed two large gaps and produced a dense map with precise marker distances integrated with gene position and orientation. Few clones were found in common in the two maps, because we used PACs and cosmids instead of BACs. In this region thought to be associated to DS congenital heart and gut diseases (Korenberg et al. 1992), only nine genes are known to date: c21–LRP (Vidal-Taboada et al. 1998); HMG14 (Petersen et al. 1990); WRB (Egeo et al. 1998); SH3BGR (Scartezzini et al. 1997); PCP4 (Chen et al. 1996; Hubert and Korenberg 1997); DSCAM (Yamakawa et al. 1998); MX2, MX1 (Horisberger et al. 1988); and TMPRSS2 (Paoloni-Giacobino 1997). DSCAM, a novel member of the immunoglobulin superfamily potentially involved in neural differentiation, spans a previous 800-kb gap in the map. No EST has been mapped in the region spanned by DSCAM, and it will be interesting to see if another gene

| Clone 1 (size in kb) | Clone 2 (size in kb) | overlaps | probability EcoRI | probability HindIII | overlap in % of clone length | overlap by sequencing (in kb) | overlap in % of clone length |
|----------------------|----------------------|----------|-------------------|---------------------|-------------------------------|-------------------------------|-------------------------------|
| 82F9 (119)           | 141B3 (118)          | /        | 5e-05             | 8e-08               | 16                            | 27                            |
| 95K20 (121)          | 279K20 (136)         | 71/36    | 2e-04             | 7e-04               | 8/16                          |                               |
| 95K20 (121)          | 141B3 (118)          | ns/no    | 1e-01             | 3e-02               | 4/2                           |                               |
| 141B3 (118)          | 31K18 (145)          | ns/no    | 1e-03             | 7e-03               | 1/7                           |                               |
| 128M19 (125)         | 105P11 (89)          | ns/no    | 4e-01             | 2e-02               | 2/1                           |                               |
| 185P11 (89)          | 160J9 (142)          | ns/no    | 4e-02             | 2e-02               | 2/1                           |                               |
| 229A2 (148)          | 160J9 (142)          | ns/no    | 1e-02             | 1e-02               | 1/1                           |                               |
| 279A2 (40)           | 160J9 (142)          | ns/no    | 1e-06             | 8e-02               | 2/2                           |                               |
| 70D24 (152)          | 228A2 (148)          | 65/94    | 1e-10             | 1e-16               | 2/2                           |                               |
| 70D24 (152)          | 241G8v (40)          | ns/no    | 2e-03             | 2e-03               | 2/2                           |                               |
| 95K20 (121)          | 160J9 (142)          | ns/no    | 2e-03             | 2e-03               | 2/2                           |                               |
| 54E1 (127)           | 7A42 (139)           | 1e-05    | 9e-06             | 1e-05               | 2/2                           |                               |
| 54G1 (40)            | 206A10 (148)         | 2e-07    | 10/16             | 1e-10               | 1e-10                         |                               |
| 54G1 (40)            | 54E4 (127)           | 1e-01    | 1e-01             | 1e-01               | 1e-01                         |                               |
| 7A42 (139)           | 160M2 (142)          | ns/no    | 1e-05             | 1e-05               | 1e-05                         |                               |
| 11A18 (142)          | 211L3 (117)          | ns/no    | 1e-01             | 1e-01               | 1e-01                         |                               |
| 211L3 (117)          | 221D19 (150)         | 7e-02    | 4e-12             | 2e-15               | 4e-12                         |                               |
| 211L3 (117)          | 198B8 (150)          | ns/no    | 4e-01             | 1e-04               | 1e-04                         |                               |
| 211L3 (117)          | 198E8 (150)          | ns/no    | 3e-01             | 1e-04               | 1e-04                         |                               |
| 211L3 (117)          | 198F12 (150)         | ns/no    | 3e-01             | 1e-04               | 1e-04                         |                               |
| 33D13 (108)          | 221D19 (150)         | ns/no    | 5e-08             | 9e-08               | 9e-08                         |                               |
| 33D13 (108)          | 221D19 (150)         | ns/no    | 1e-02             | 1e-02               | 1e-02                         |                               |
| 33D13 (108)          | 247E2 (151)          | ns/no    | 1e-02             | 1e-02               | 1e-02                         |                               |
| 33D13 (108)          | 247F2 (151)          | 1e-02    | 8e-05             | 7e-05               | 7e-05                         |                               |
| 33D13 (108)          | 247G2 (151)          | 1e-02    | 8e-05             | 7e-05               | 7e-05                         |                               |
| 251L6 (99)           | 146B4 (103)          | 2e-03    | 1e-03             | 1e-03               | 1e-03                         |                               |
| 251L6 (99)           | 146B4 (103)          | 1e-03    | 1e-03             | 1e-03               | 1e-03                         |                               |
| 251L6 (99)           | 146B4 (103)          | 1e-03    | 1e-03             | 1e-03               | 1e-03                         |                               |
| 251L6 (99)           | 146B4 (103)          | 1e-03    | 1e-03             | 1e-03               | 1e-03                         |                               |
| 251L6 (99)           | 146B4 (103)          | 1e-03    | 1e-03             | 1e-03               | 1e-03                         |                               |
| 105K13 (150)         | 269A4 (100)          | ns/no    | 1e-05             | 3e-11               | 3e-11                         |                               |
| 105K13 (150)         | 269A4 (100)          | ns/no    | 1e-05             | 3e-11               | 3e-11                         |                               |
| 105K13 (150)         | 269A4 (100)          | ns/no    | 1e-05             | 3e-11               | 3e-11                         |                               |
| 105K13 (150)         | 269A4 (100)          | ns/no    | 1e-05             | 3e-11               | 3e-11                         |                               |
| 105K13 (150)         | 269A4 (100)          | ns/no    | 1e-05             | 3e-11               | 3e-11                         |                               |
| 105K13 (150)         | 269A4 (100)          | ns/no    | 1e-05             | 3e-11               | 3e-11                         |                               |
| 105K13 (150)         | 269A4 (100)          | ns/no    | 1e-05             | 3e-11               | 3e-11                         |                               |
will be predicted there by genomic sequencing. We have isolated 14 anonymous potential transcripts between S3 and MX. It is difficult to estimate how many genes they represent since some may cluster within the same gene. Interestingly, a partial cDNA from TU188 (accession no. AJ003597) shows 60% homology with β-1,3-galactosyltransferases. β-Galactosyltransferases belong to a gene family of wide structural diversity, playing a role in cell adhesion and cell proliferation processes (Shur 1993; Hathaway and Shur 1996). These gene signatures will assist experimental confirmation of novel gene models predicted in silico from genomic sequence. We can estimate that the S3–TMPRSS2 contig contains ~20 genes, thus representing a gene density of 1/150 kb.

Large-scale genomic sequencing projects rely on the concomitant development of bacterial clone-based maps. Establishing the basic clone resource is often a trade-off between coverage and contiguity. We have shown here that nonisotopic hybridization technology can be successfully applied to build maps fulfilling these criteria. To the best of our knowledge, the contig shown in this study is gapless between S3 and TMPRSS2, although we cannot rule out small rearrangements such as internal deletions of the chosen clones. The corresponding sequence is freely available (http://genome.imb-jena.de/ and http://seq.mpimg-berlin-dahlem.mpg.de). Importantly, the sequenced map will provide essential molecular tools for diagnostic purposes, positional candidate gene approach, and in-depth analysis of all genes and associated regulatory elements in the DS-CHD region.

### METHODS

#### Nonradioactive Screening of Arrayed Genomic Libraries

Three cosmid libraries have been used, constructed from flow-sorted chromosome 21. The LL21CO2 "Q" library constructed at the Lawrence Livermore National Laboratory contains 10,464 clones, representing a sevenfold chromosome coverage. The ICRFc102 and ICRFc103 libraries were constructed at the Imperial Cancer Research Foundation (ICRF) (Nizetic et al. 1991) and contain together 16,896 clones representing 11 equivalents of chromosome 21.

We have used a part of the PAC RPCI-1 library (Ioannou et al. 1994), representing 2.5 times human genome equivalents.

All libraries were gridded at high density onto Hybond N+ nylon membranes (Amersham). Clones were spotted in duplicate using a 5 × 5 pattern with guide dots to facilitate scoring of positives. This spotting pattern allows more than 55,000 clones to be spotted onto one 22 × 22-cm nylon membrane. Filters were laminated onto a plastic support (Bancroft et al. 1997) and bar-coded to facilitate data tracking and high throughput handling.

All probes were labeled with digoxigenin (DIG-11–dUTP) (Boehringer Mannheim). STSs and cDNAs were labeled by incorporating DIG-11–dUTP by PCR substituting part of the dTTP in the reaction (0.2 mM dATP, dCTP, and dGTP; 0.19 mM dTTP/0.01 mM DIG-11–dUTP). Riboprobes were prepared from 1–2 µg of template DNA and synthesized by in vitro transcription using SP6 or T7 RNA polymerases according to the manufacturer recommendations (Boehringer Mannheim). To block repetitive sequences, probes were competed with 100 µg of sheared human placental DNA (Sigma) at 65°C for 1.5 hr.

DIG-labeled probes were denatured for 10 min and diluted in 2.5 ml of Church buffer for each filter (Church buffer

### Table 4. (Continued)

| Clone No. | Size (kb) | FPC Value | E. coli | Hin dIII | Eco RI | Overlap (kb) | Probability |
|-----------|----------|-----------|---------|----------|--------|-------------|-------------|
| 182T14    | 147      | 89/59     | 4e-19   | 8e-12    | 56/31  |             |             |
| 182T16    | 147      | nso/58    | 5e-04   | 4e-12    | /26    |             |             |
| 182T27    | 147      | 265b (26') | nso/46  | 1e-01    | 4e-10  | /19         |             |
| 182T28    | 147      | 355b (150) | nso/nso | 3e-01    | 7e-04  |             |             |
| 182T29    | 147      | 265b (125) | nso/22  | 5e-02    | 5e-13  | /11         |             |
|           |          | Q14b (10') | nso/8e  | 9e-05    | 8e-05  |             |             |
|           |          | 265b (137) | nso/52  | 8e-10    | 5e-13  | 16/25       |             |
|           |          | 265b (105) | nso/40  | 4e-15    | 30/30  |             |             |
|           |          | 265b (137) | Q15d (40) | 5e-02   | 4e-02  |             |             |
|           |          | 265b (40)  | Q15d (44) | nso/3e  | 9e-01  |             |             |

Clone sizes (kb) indicated in brackets were generally estimated by PFGE, or by the mean FPC value form EcoRI and HindIII patterns (1) by the FPC value from EcoRI pattern (2) or from genomic sequencing (3). Overlaps (in kb) between clones estimated by PFGE after EcoRI or HindIII digestion are shown with the corresponding probability value: (nso) no significant overlap; (/) no FPC data available. When available, overlaps calculated from genomic sequence are given (in kb). Clone to clone overlaps were also calculated in percent of the actual length spanned by a given pair of clones. Cosmids are prefixed with a Q and BACs with a B. The sequenced clones are indicated with an x in the first column. Sequences are available at http://genome.imb-jena.de (Note: cosmids are displayed in 384-well format on this server; Q73F12 = 19K23, Q41G8 = M1531, Q7C4 = 2N7, Q82G10 = 21M20, and Q7B67-20C22) or at http://seq.mpimg-berlin-dahlem.mpg.de (identical names).
consists of 0.25 M Na₂HPO₄ at pH 7.2, 1.25 mM EDTA, and 5% SDS and is warmed to 65°C before hybridization). The probe solution was evenly sprayed onto each filter using an air-spray gun (air pressure, 3.5-4 bar), and laminated filters were stacked on top of each other in a plastic bag. Generally, we placed 40 large filters per bag, and all of the washing solution volumes are given for 40 filters. Hybridizations were performed overnight at 65°C. Filters were washed in 2 liters of 20 mM Na₂HPO₄, 0.1% SDS, for 15 min at room temperature, followed by a second wash for 20 min in the same solution at 65°C.

Detection was performed using the Attaphos system (Maier et al. 1994). Prior to detection, filters were washed in a 2-liter blocking solution (5% milk powder/PBS) solution for 45 min at room temperature. Each filter was sprayed with 2.5 ml of anti-digoxigenin–alkaline phosphatase antibody solution (Boehringer Mannheim; diluted 1:5000 in 5% milk powder/PBS), using the air-spray gun. Excess of antibody was removed by washing the filters two times each for 30 min in 2-liter volumes of PBS at room temperature. Filters were washed two times in 2-liter volumes of 0.1 M Tris, 1 mM MgCl₂ (pH 9.5) at room temperature for 10 min. Then, filters were sprayed with a 2.5-mI solution containing the fluorescent substrate Attaphos (JBL Scientific) (0.08 M Tris/0.8 mM MgCl₂ at pH 9.5, 0.48 M diethanolamine/0.046 mM MgCl₂/0.001% Na₂S₂O₃, and 1 mM Attaphos). After incubation at 37°C for 4 hr, positive signals were visualized under long-wave UV light. Images were captured with a charge-coupled device (CCD) camera and saved as a TIFF format file.

Images were processed and scored semiautomatically using the Xdigitize program developed in our department. Xdigitize converts the position of positives on the gridded filter to the actual clone name in the library. Data were transferred automatically to a dedicated database, and analysis proceeded with the well2clone program for primary contig assembly (Mott et al. 1993). Well2clone is a dedicated suite of programs written for building contigs from hybridization data. It serves a dual purpose, providing a database managing all hybridization data (listing probes and matching positives) and allowing the building of contigs using the related functions probeorder or reorder. well2clone is freely available by ftp at (ftp://ftp.mpimg-berlin-dahlem.mpg.de/pub/lehrach/users/andy/icrf_contigv2.6.tar.Z). Xdigitize is freely available upon request.

Restriction Digest Fingerprinting

DNAs (8 µg/reaction) were digested with either EcoRI or HindIII restriction enzymes. DNA fragments were separated onto agarose gels (25.5 cm, cast in 0.8% agarose), electrophoresed in 1 X TAE buffer for 24 hr at 1.6 V/cm. DNA staining was performed with Vistra Green (0.01%) (Amersham), and images were captured with the Fluor-Imager 575 (Molecular Dynamics). Gel images were analyzed with the program Image3 (Sulston et al. 1989). The data from Image3 was exported to the FPC program (Soderlund et al. 1997). FPC is an interactive program used for building contigs from fingerprinted clones. Analysis was performed setting the following parameters: tolerance = 7, cutoff = 10⁻⁷, diff = 0.3, diffbury = 0.1, minimal bands = 3, and minimal ends = 8. Restriction patterns were compared for shared fragments, and overlapping clones were clustered automatically. In all cases, overlaps were verified manually by the user.

For Southern blotting, DNAs were transferred onto Hybond N+ membranes (Amersham) using standard conditions. Radioactive riboprobes were labeled with 50 µCi [α-³²P]CTP. Radioactive riboprobes were competed as for DIG-labeled probes. Hybridizations were performed in (5 X SSPE, 5 X Denhardt’s, 50% formamide, 1% SDS, and 5 µg/ml salmon sperm DNA) overnight at 42°C. Filters were washed to (2 X SSC/0.1% SDS) at 65°C. Filters were exposed onto X-ray films (Kodak, X-OMAT-AR).

PFGE

PAC DNAs (1–2 µg) were digested with NofI restriction enzyme (Biolabs) for 2 hr at 37°C in the appropriate buffer. PFGE was performed in 0.5 X TBE and 1% agarose gels (Seakem GTG; FMC), using the CHEF-DRII apparatus (Bio-Rad). Electrophoresis was performed at 6 V/cm for 20 hr at 14°C with a 4- to 20-sec ramp. DNA markers were long-range-ladder (multimers of 48.5 kb; Bio-Rad) and mid-range χ DNA (monocut mix; Biolabs).

STS Content

PCRs were performed in PCR buffer [10 mM Tris-HCl at pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, and 0.001% (wt/vol) gelatin], with 10 pmoles of each primer, 0.2 mM dNTPs, and 1.25 unit of Taq polymerase (AmpliTaq) (Perkin-Elmer Cetus) in a total volume of 25 µl. Amplifications were performed with a denaturation step at 94°C for 3 min, followed by 35 cycles of amplification (94°C for 45 sec, annealing at primer-specific temperatures for 30 sec and extension at 72°C for 1 min) and a final extension at 72°C for 4 min.

FISH

Metaphase chromosomes were prepared from human peripheral blood lymphocytes. Standard FISH protocols were followed (Ward et al.1995). Briefly, the slides were treated with 100 µg/ml RNase A in 2 X SSC at 37°C for at least 30 min and with 0.01% pepsin in 10 mM HCl at 37°C for 10 min and then dehydrated in an ethanol series (70%, 85%, and 100%). Slides were denatured at 95°C in 70% formamide and 2 X SSC (pH 7.0) and again dehydrated in a cold alcohol series. Probes were labeled by standard nick translation procedures with either biotin-16-dUTP or digoxigenin-11-dUTP (Boehringer Mannheim). Labeled DNA was coprecipitated with human Cot-1 competitor DNA (GIBCO-BRL) and herring sperm carrier DNA and redissolved in 50% formamide, 10% dextran sulfate, and 2 X SSC. After a 10-min denaturation at 80°C, 10 µl of hybridization mixture was applied to each slide and sealed under a coverslip. Slides were left to hybridize in a moist chamber at 37°C for 1–3 days. Slides were washed to 0.1 X SSC for 5 min at 60°C. Biotinylated probes were detected by fluorescein isothiocyanate (FITC)-conjugated avidin (Vector) and digoxigenated probes by Cy3-conjugated anti-digoxigenin antibody (Dianova). Chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI). Images were taken with a Zeiss epifluorescence microscope equipped with a thermoelectronically cooled CCD camera (Photometrics CH250).

NOTE

Clone names that are not given in Figure 2 owing to space limitations are freely available upon request and on the primary database of the RZPD server (http://www.rzpd.de/).
ACKNOWLEDGMENTS

We thank the RZPD Resource Center for supplying filters and clones. Jörn Steiger and Beatrix Roehrdanz are thanked for excellent technical assistance. We thank Huw Griffith for the Xdigitize program, Dr. Juliane Ramser and Dr. Stefan Tau din for sharing genomic sequence information, Dr. Thomas Haaf for providing the FISH facilities, Dr Peter de Jong for the LL21NCCO2 "Q" and PAC libraries, and Dr. Julie Korenberg for supplying BACs 28F9 and 291B3. This work was supported by the Deutsches Humangenomprojekt (DMBF grant 01K W 9608).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Bancroft, D.R., J.K. O'Brien, A. Guerasim ova, and H. Lehrach. 1997. Simplified handling of high density gen filters using rigid plastic laminates. Nucleic Acids Res. 25: 4160–4161.

Chen, H.M., C. Bouras, and S.E. Antonarakis. 1996. Cloning of the cDNA for a human homolog of the rat pep-19 gene and mapping to chromosome 21q22.2-2q22.3. Hum. Genet. 98: 672–677.

Chumakov, I., P. Rigault, S. Guillou, P. Oug en, A. Billaut, G. Guasco ni, P. Gervy, I. LeGall, P. Soul arue, L. Grinas et al. 1992. Continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359: 380–387.

Church, D.M., C.J. Stotler, J.L. Rutter, J.R. Murrell, J.A. Trofatter, and A.J. Buckler. 1994. Isolation of genes from complex sources of tiling path construction for chromosome 21q. Genomics 18: 735–745.

Eigeo, A., F. Sotgia, P. Arrigo, T. Nizetic, A. Hernandez, D. and E.M.C. Fisher. 1996. Down syndrome genetics—Unravelling a multifactorial disorder. Hum. Mol. Genet. 5: 1411–1416.

Hoheisel, J.D., E. Maier, R. Mott, L. McCarthy, A.V. Grigoriev, L.C. Schalkwijk, D. Nizetic, F. Francis, and H. Lehrach. 1993. High resolution cosm id and P1 maps spanning the 14 Mb genome of the fission yeast S. pombe. Cell 73: 109–120.

Horisberger, M.A., M. Wathelet, J. Szpirer, C. Szpirer, Q. Islam, G. Levan, G. Hucz, and J. Content. 1998. cDNA cloning and assignment to chromosome 21 of IFI-78K gene, the human equivalent of murine Mx gene. Somat. Cell Mol. Genet. 14: 123–131.

Hubert, R.S. and J.R. Korenberg. 1997. PCP4 maps between D21S345 and P31P105F6 on chromosome 21q22.2 — q22.3. Cytogenet. Cell. Genet. 76: 44–45.

Hubert, R.S., S. Mitchell, X.N. Chen, E. Kmek j, C. Gadomski, Z. Sun, D. Noya, U.J. Kim, C. Chen, H. Shizuya et al. 1997. BAC and PAC contigs covering 3.5 Mb of the Down syndrome congenital heart disease region between D21S55 and MX1 on chromosome 21. Genomics 41: 218–226.

Ichikawa, H., K. Shinimi, A. Saito, D.N. Wang, R. Oli va, H. Koba yashi, Y. Kaneko, H. Miyoshi, C.L. Smith, C.R. Cantor et al. 1992. Long-distance restriction mapping of the proximal long arm of human chromosome 21 with Not I linking clones. Proc. Nat l. Acad. Sci. 89: 23–27.

Ioannou, P.A., C.T. Amemiya, J. Garnes, P.M. Kroisel, H. Shizuya, C. Chen, M.A. Batzer, and P.J. de Jong. 1994. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nature Genet. 6: 84–89.

Korenberg, J.R., C. Bradley, and C.M. Dist echte. 1992. Down syndrome: Molecular mapping of the congenital heart disease and duodenal stenosis. Am. J. Hum. Genet. 50: 294–302.

Korenberg, J.R., X.N. Chen, R. Schipper, Z. Sun, R. Gonsky, S. Gerwehr, N. Carpenter, C. Daumer, P. Dignan, C. Dist echte et al. 1994. Down syndrome phenotypes: The consequences of chromosomal imbalance. Proc. Natl. Acad. Sci. 91: 4997–5001.

Korn, B., Z. Sedlacek, A. Manca, P. Kioschis, D. Konecki, H. Lehrach, and A. Poustka. 1992. A strategy for the selection of transcribed sequences in the Xq28 region. Hum. Mol. Genet. 1: 235–242.

Lauer, P., N.C. Meyer, C.E. Prass, S.M. Starnes, R.K. Wolff, and A. Guirke. 1997. Clone-contig and STS maps of the hereditary hemochromatosis region on human chromosome 6p21.3-p22. Genome Res. 7: 457–470.

Maier, E., H.R. Crollius, and H. Lehrach. 1994. Hybridisation techniques on gridded high density DNA and in situ colony filters based on fluorescence detection. Nucleic Acids Res. 22: 3423–3424.

Mara, M.A., T.A. Kucaba, N.L. Dietrich, E.D. Green, B. Browstein, R.K. Wilson, K.M. McDonal, W.H. LaDeana, J.D. McPherson, and R.H. Waterston. 1997. High throughput fingerprint analysis of large-insert clones. Genome Res. 7: 1072–1084.

Mott, R. A. Grigoriev, E. Maier, J. Hoheisel, and H. Lehrach. 1993. Algorithms and software tools for ordering clone libraries: Application to the mapping of the genome of Schizosaccharomyces pombe. Nucleic Acids Res. 21: 1965–1974.

Niederfuehr, A., H. Hummerich, B. Gawin, S. Boyle, P.F.R. Little, and M. Gessler. 1998. A sequence-ready 3 Mb PAC contig covering 16 breakpoints of the Wilm's Tumor/Aniridia region of human chromosome 11p13. Genomics 53: 153–163.

Nizetic, D., G. Zehetner, A.P. Monaco, L. Gellen, B.D. Young, and H. Lehrach. 1991. Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: Their potential use as reference libraries. Proc. Natl. Acad. Sci. 88: 3233–3237.

Nizetic, D., L. Gellen, R.M. Hamvas, R. Mott, A. Grigoriev, R. Vatcheva, G. Zehetner, M.L. Yaso, A. Dutiauxa, C. Lopes et al. 1994. An integrated YAC-overlap and "cosmid-pocket" map of the human chromosome 21. Hum. Mol. Genet. 3: 759–770.

Ohira, M., H. Ichikawa, E. Suzuki, M. Iwaki, K. Suzuki, F. Saitoohara, T. Ikeuchi, I. Chumakov, H. Tanashahi, K. Tashiro et al. 1996. A 1.6-Mb P1-based physical map of the down syndrome region on chromosome 21. Genomics 33: 65–74.
Ohira, M., N. Seki, T. Nagase, E. Suzuki, N. Nomura, O. Ohara, M. Hattori, Y. Sakaki, T. Eki, Y. Murakami et al. 1997. Gene identification in 1.6-Mb region of the Down syndrome region on chromosome 21. *Genome Res.* **7**: 47–58.

Osoegawa, K., R. Susukida, S. Okano, J. Kudoh, S. Minoshima, N. Shimizu, P.J. Dejong, J. Geert, J. Ives, H. Lehrach et al. 1996. An integrated map with cosmid/pac contigs of a 4-Mb Down syndrome critical region. *Genomics* **32**: 375–387.

Paoloni-Giacobino, A., H. Chen, M.C. Peitsch, C. Rossier, and S.E. Antonarakis. 1997. Cloning of the TMPRSS2 gene, which encodes a novel serine protease with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3. *Genomics* **44**: 309–320.

Parimoo, S., S.R. Patanjali, H. Shukla, D.D. Chaplin, and S.M. Weissman. 1991. cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci.* **88**: 9623–9627.

Petersen, M.B., E.P. Economou, S.A. Slaugenhaupt, A. Chakravarti, and S.E. Antonarakis. 1990. Linkage analysis of the human HMG14 gene on chromosome 21 using a GT dinucleotide repeat as polymorphic marker. *Genomics* **7**: 136–138.

Rahmani, Z., J.L. Bioulé, N. Creau-Goldberg, P.C. Watkins, J.F. Mattei, M. Poissonnier, M. Prieur, Z. Chettouh, A. Nicole, A. Auriol et al. 1989. Critical role of the D21S55 region on chromosome 21 in the pathogenesis of Down syndrome. *Proc. Natl. Acad. Sci.* **86**: 5958–5962.

Scartezzini, P., A. Egeo, S. Colella, P. Fumagalli, P. Arrigo, D. Nizetic, R. Taramelli, and A. Rasore-Quartino. 1997. Cloning a new human gene from chromosome 21q22.3 encoding a glutamic acid-rich protein expressed in heart and skeletal muscle. *Hum. Genet.* **99**: 387–392.

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, P. Rodriguez-Tome, A. Aggarwal, E. Bajorek et al. 1996. Genome maps. The human transcript map. *Science* **274**: 547–558.

Shur, B.D. 1993. Glycosyltransferases as cell adhesion molecules. *Curr. Opin. Cell Biol.* **5**: 854–863.

Soderlund, C., I. Longden, and R. Mott. 1997. FPC: A system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.* **13**: 523–535.

Soeda, E., D.X. Hou, K. Osoegawa, Y. Atsuchi, T. Yamagata, T. Shimokawa, H. Kishida, E. Soeda, S. Okano, I. Chumakov et al. 1995. Cosmid assembly and anchoring to human chromosome 21. *Genomics* **25**: 73–84.

Sulston, J., F. Mallett, R. Durbin, and T. Horsnell. 1989. Image analysis of restriction enzyme fingerprint autoradiograms. *Comput. Appl. Biosci.* **5**: 101–106.

Vidal-Taboada, J.M., S. Sanz, A. Egeo, P. Scartezzini, and R. Oliva. 1998. Identification and characterization of a new gene from human chromosome 21 between markers D21S343 and D21S268 encoding a leucine-rich protein. *Biochem. Biophys. Res. Commun.* **250**: 547–554.

Ward, D.C., A. Boyle, and T. Haaf. 1995. Fluorescence in situ hybridization techniques. Metaphase chromosomes, interphase nuclei, and extended chromatin fibers. In *Human chromosomes. Principles and techniques* (ed. R.S. Verma and A. Babu), pp. 184–192. MacGraw-Hill, New York, NY.

Yamakawa, K., M.A. Haendel, R. Hubert, X.-N. Chen, G.E. Lyons, and J.R. Korenberg. 1998. DSCAM: A novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. *Hum. Mol. Genet.* **7**: 227–237.

Yaspo, M.L., L. Gellen, R. Mott, B. Korn, D. Nizetic, A.M. Poustka, and H. Lehrach. 1995. Model for a transcript map of human chromosome 21—Isolation of new coding sequences from exon and enriched cDNA libraries. *Hum. Mol. Genet.* **4**: 1291–1304.

Received April 17, 1998; accepted in revised form January 29, 1999.