
Recombination-Based Screening for Genes on Chromosome 21

Gordon D. Stewart and David M. Kurnit

Howard Hughes Medical Institute, Department of Pediatrics (G.D.S., D.M.K.) and Department of Human Genetics (D.M.K.), University of Michigan, Ann Arbor, Michigan

In previous work, a reasonably large number of genes was found to be on chromosome 21. This frequency demands the creation of new techniques to investigate the transcription of this small human autosome, and to relate these findings to the phenotype of Down syndrome. Here we describe the elaboration of new vectors and hosts, which in conjunction with flow-sorted cosmid libraries of chromosome 21 and genic (cDNA) libraries of relevant human tissues, will permit us to examine transcription of chromosome 21.

KEY WORDS: cDNA library, bacteriophage λ , chromosome 21

INTRODUCTION

The isolation of genes encoded by chromosome 21 remains a major research goal for the elucidation of Down syndrome (DS). Not surprisingly, attempts have been made to explain the clinical results of trisomy 21 on the basis of the known genes encoded by this chromosome. However, the bulk of genes encoded by this autosome have neither been identified nor isolated to date. Here we describe methods to isolate gene sequences on chromosome 21 that are transcribed in particular tissues, paralleling the genomic effort that will occur on this smallest human autosome.

RESULTS AND DISCUSSION

The older methods we used relied on hybridization to isolate cDNA clones encoded by chromosome 21. In our first approach, a cDNA library representing the expressed genes of a mouse-human somatic cell hybrid (WA17) was screened with total human DNA to identify human chromosome 21-specific cDNAs [Neve et al., 1986]. Although this technique successfully detected 3

expressed sequences on human chromosome 21, it did not detect the bulk of sequences encoded by that chromosome, consistent with work which demonstrated that most human genes in such hybrids are not transcribed. In the second approach [Neve et al., 1986; Neve et al., unpublished], previously characterized single-copy anonymous genomic fragments on chromosome 21 were used as probes to retrieve homologous coding sequences from a human fetal brain library. Again, the technique worked, resulting in the isolation of new transcribed sequences on chromosome 21. Taken together, these approaches demonstrated that there are likely to be many more coding sequences on chromosome 21 than have been identified as yet by serendipity. In fact, up to one-half of the chromosome 21 genomic inserts we have analyzed in this manner have hybridized to brain cDNA. Since our sequencing studies on the cDNA clones isolated to date [McCann et al., unpublished] have not detected previously known sequences, most of the transcribed sequences on chromosome 21 have yet to be described. Thus, we had to devise a novel strategy enabling us to detect transcribed sequences on chromosome 21.

Our newer method centers on the availability of the recombination-based assay elaborated by Seed [1983]. Unfortunately, the vectors and hosts described in this 1983 paper are no longer useful. This has mandated the design of new hosts and vectors. We describe here the use of such hosts and vectors, and their application to chromosome 21.

In the original assay designed by Seed [1983], a recombinant library was constructed using ultraviolet (UV)-irradiated packaging extracts and was screened with a segment of DNA cloned in a plasmid that carried the prokaryotic selectable marker, *supF*. Two aspects of that protocol are no longer applicable:

1. UV-irradiated packaging extracts, which were required initially for safety reasons, have been supplanted by non-UV-irradiated extracts. While the latter are more efficient and will not be replaced, the libraries so constructed have unacceptably high background frequencies (10^{-6} to 10^{-7}) due to library construction in rare phages that lost their double amber mutations by recombination between vector phage and phage DNA in the packaging extracts.

Received for publication May 22, 1989; revision received August 30, 1989.

Address reprint requests to Gordon D. Stewart, Ph.D., Howard Hughes Medical Institute, University of Michigan Medical Center, 1150 W. Medical Center Drive, MSRB I, Room 3520, Ann Arbor, MI 48109-0650.

2. Once the forward reaction (recombination to acquire the plasmid + incorporated sequences) occurred, it was not feasible to reverse the reaction. The ability to reverse the reaction is important on two accounts: it permits the mapping of library sequences without the confounding effects of inserts that recombined in, and the frequency of reversal is used to identify and discard nonhomologous recombination events. A low (ca. 10^{-9}) frequency of reversal is associated with nonhomologous recombination, and can be used to diagnose these unwanted events.

Thus, to use the recombination-based assay on chromosome 21, we made multiple improvements to the original procedure:

1. Construction of a new host to eliminate background in the forward reaction.
2. Construction of a new vector and host to permit counterselection (and hence recombination-based excision) of plasmids bearing supF in the reverse reaction.

In the recombination-based assay, a complex library is searched rapidly using in vivo screening for the presence of a given sequence. To search for genes encoded by human chromosome 21, we screen a genic (cDNA) library with a genomic sequence from human chromosome 21. The screen works as follows. The genomic sequence is cloned into a specialized plasmid vector that carries the bacterial suppressor (supF) gene. This mutant tRNA gene inserts a tyrosine at one of the stop codons (UAG, the "amber" codon) in the genetic code. In this manner, it is possible to bypass a stop (UAG) signal in the genetic code. We have constructed a specialized host, DK21 [Kurnit and Seed, 1990] that selects for phage carrying such a suppressor gene. (DK21 carries an amber mutation in the *dnaB* gene that is required for phage, but not bacterial, growth in the presence of the *ban* gene from phage P1.) To screen for genes on chromosome 21, we carry out the following:

1. As stated previously, a genomic sequence from human chromosome 21 is mapped subregionally on this chromosome, and inserted into a specialized plasmid with supF. Alternatively, the insertion of chromosome 21 DNA fragments into such a plasmid vector can be done in a "shotgun" fashion preceding subregional mapping.

2. A complex cDNA library in a bacteriophage, corresponding to sequences expressed in a given tissue or cell line (e.g., brain), is plated on an overnight culture of bacteria carrying the plasmid with the genomic chromosome 21 sequence to be analyzed. To cover the cDNA library, 10^6 to 10^7 recombinants spanning the complexity of the cDNA library are plated. The resulting plate is lysed, and eluted into phage buffer.

3. Incorporation of the plasmid into the bacteriophage by homologous recombination (between a genomic chromosome 21-DNA sequence cloned in the plasmid and a genic sequence expressed in the cDNA library) is entrapped by selecting on DK21 for phages

that have acquired supF. As a result, the assay yields sequences on chromosome 21 that are transcribed, i.e., sequences on chromosome 21 that encode genes present in a given cDNA library.

4. To eliminate the genomic sequence on chromosome 21 from the genic (cDNA) sequence, we must reverse the recombination reaction detailed above. To perform this reversal, supF (which was selected for previously) must be made lethal to the host. Currently, such lethality is available only when supF is on a plasmid vector. This counterselection against supF, which we term "galactocide," relies on the fact that galactose is lethal to galK⁺ galE⁻ hosts. Using a galK amber mutation and a galE point mutation, supF becomes lethal as it suppresses the galK, but not the galE, mutation; the accumulation of UDP-galactose (mediated by galK that cannot be processed further in the absence of galE) is lethal to the bacterial cell.

Thus, we perform our recombination-based assay against cDNA libraries cloned in a phasmid vector (Sumo15A [Kurachi et al., 1989]), which propagates as either a phage or plasmid, depending on the properties of the bacteria in which it is propagating. To select for supF, we plate cDNA libraries on a host, DK21 [Kurnit and Seed, 1990] that selects for the phage state of Sumo15A; to select against supF, we plate a selected recombinant on a host, DK37 [Kurnit, 1989], in which supF is lethal to the plasmid state of Sumo15A.

Construction of a host to eliminate background in the forward recombination reaction. Selection for recombinants was performed previously by plating a double amber phage library on a Su⁰ host, and then screening for phages that had integrated supF and could propagate on such a host. Unfortunately, the creation of rare (10^{-6} to 10^{-7}) am⁺ phages during library construction vitiated the selection scheme which required the screening of 10^8 to 10^9 phages from a cDNA library to accomplish the screening of a cDNA library containing a complexity of 10^5 to 10^6 (since only 10^{-2} to 10^{-3} homologous phages undergo recombination). This necessitated a different screening strategy that required a further selection for phages carrying supF. After multiple attempts, we were able to construct a host that indeed required supF acquisition for phage propagation. This host, DK21 [*lacZam dnaBam266* (λ *imm21 b515 b519 nin5 attP P1 ban*)] [Kurnit and Seed, 1990; shown in Fig. 1], utilizes the *dnaBam* gene to select for λ phages carrying supF (acquired via recombination). Although the DK21 host can replicate using the *P1 ban* gene in place of the *E. coli dnaB* primase gene, phages lacking supF cannot so replicate. Only phages carrying supF can suppress the *dnaBam* mutation, allowing these phages to propagate on the *dnaBam* host. This selection works at the 10^{-8} level. The recombinants can then be assayed further using the *lacZam* gene: true recombinant phages will yield blue plaques (due to suppression by supF of the *lacZam* allele), while nonrecombinants will not suppress the *lacZam* allele and will yield clear plaques. Further, the *dnaBam* selection is additive to selection based on using amber vectors. Although the selection scheme using amber vectors is not tight

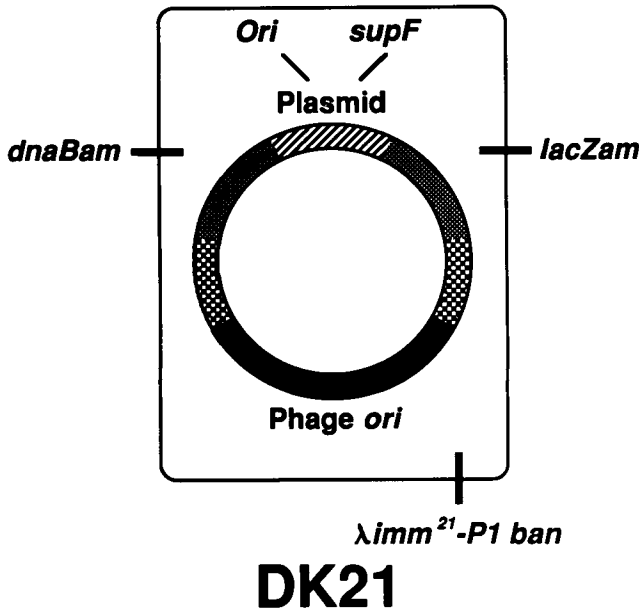


Fig. 1. DK21: Selection for phages carrying supF. DK21 has the following genotype: *lacZ_{YA536}(am) dnaB266(am) str^r (λ imm²¹ b515 b519 nin5 attP P1 ban)*. As outlined in the text, only Sumo15A phages carrying supF can plate on DK21, yielding blue plaques on plates with IPTG (isopropyl-β-D-thiogalactopyranoside) and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside).

enough for library screening (10^{-6} to 10^{-7}), combination with the *dnaBam* selection scheme (10^{-8}) yields selection (10^{-14} to 10^{-15}) which is sufficiently tight that screening detects only phages that incorporated supF (for overall scheme, see Fig. 2).

Construction of a new vector and host to permit counterselection (and hence excision) of plasmids bearing supF. Following selection on DK21 of cDNA-bearing phages that have integrated genomic sequences from chromosome 21, we need to reverse the recombination event (Fig. 2) for 2 reasons. First, recombination that has occurred because of true homology (10^{-3} rate of recognition times 10^{-5} inverse of complexity of a cDNA library = 10^{-8}) will reverse at a rate equivalent to the high rate of insertion (ca. 10^{-3}), whereas recombination that has occurred due to random integration (10^{-9}) will reverse at the same low rate of insertion (ca. 10^{-9}) [Ikeda et al., 1982; Marvo et al., 1983]. Thus, although the forward rates of legitimate and illegitimate recombination are similar, the reverse rates are dissimilar by 6 orders of magnitude, allowing one to distinguish legitimate from illegitimate recombination [Kurachi et al., 1989]. Second, the reversal event allows one to eliminate chromosome 21 genomic sequences from the phages, enabling one to map a cDNA retrieved by recombination without interference from the genomic DNA that has previously been mapped to chromosome 21.

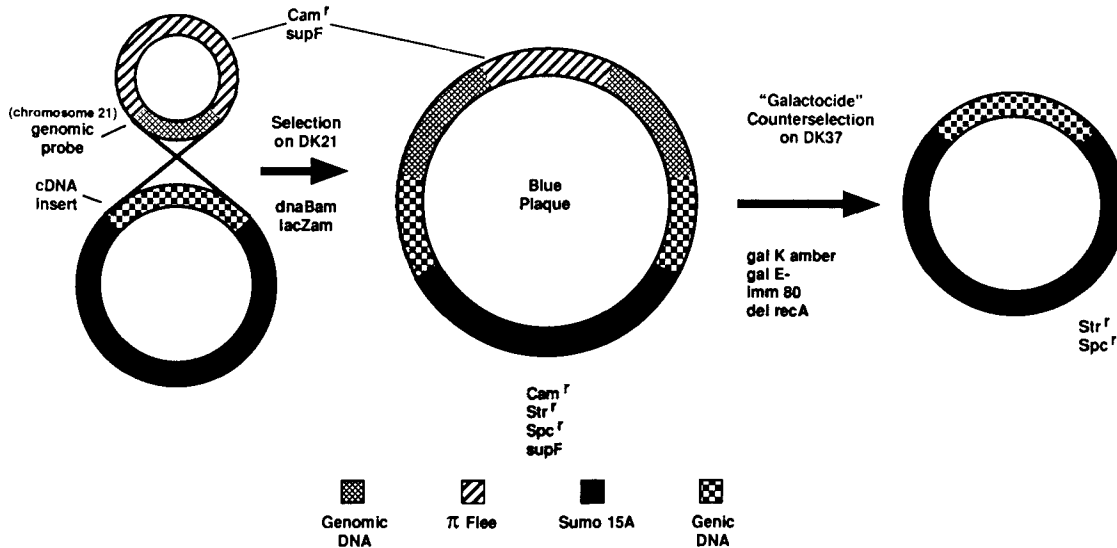


Fig. 2. Recombination-based assay (adapted from Kurachi et al. [1989]). The case is shown where the genomic chromosome 21 insert in the small plasmid πFlee matches the cDNA insert in the phasmid vector Sumo15A. Integration of the πFlee plasmid into the Sumo15A phasmid is mediated by recombination based on homology of a region of the inserts cloned in πFlee and Sumo15A. Following homology-mediated integration, the chimeric Sumo15A phage carries supF, and can therefore plate on the host DK21: supF suppresses the amber mutations in the cell (*dnaBam*) and the phage (*Aam Bam*). The resulting phage plaque is blue in the presence of IPTG and X-gal (Fig. 1), due to suppression by supF of the *lacZam* allele in DK21. Counterselection against supF^r carried in Sumo15A is accomplished by plating the chimeric phasmid on a mixed lawn of LE392 (*supF⁰ recA⁻*) and DK37 (*sup⁰ recA⁻ immφ80 galKam galE⁻*). Following propagation as a phage through LE392, a small number of phasmids lose supF by homologous recombination between the genomic insert originally carried by πFlee and the genic insert originally carried by Sumo15A. Sumo15A phasmids that have lost supF by recombination can then propagate as pSC101-driven plasmids conferring streptomycin and spectinomycin resistance in DK37 on galactose plates; in this host, the φ80 immunity gene product prevents lytic propagation of φ80-derived phages (such as Sumo15A), and the *galKam.galE⁻* genotype is lethal to phasmids bearing supF. Forward recombination with successful plating as a phage on DK21, followed by reversal of the recombination event with successful plating as a plasmid on DK37 is shown.

For these purposes, we developed the phasmid cloning vector, Sumo15A, and the *E. coli* host, DK37. The Sumo15A cloning vector was constructed by replacing the nonessential 2 kb *EcoRI-HindIII* fragment of λ Charon15A [Blattner et al., 1977] with the 4 kb *EcoRI-HindIII* fragment of pGB2 that spans all but the polylinker of this pSC101-derived plasmid, and that contains the streptomycin and spectinomycin resistance genes [Churchward et al., 1984]. The resulting phasmid, Sumo15A [Kurachi et al., 1989], has the following properties:

1. It has the replicon and red gene from Charon15A.
2. Like Charon15A, it can be repressed by the immunity gene product, ϕ 80, so that no lytic growth occurs in a host expressing ϕ 80 immunity.
3. Upon repression of the phage replicon by ϕ 80 immunity, Sumo15A will propagate as a streptomycin- and spectinomycin-resistant plasmid driven by the pSC101-replicon embedded in pGB2, and hence Sumo15A.

Efficient positive selection for supF occurs because of the presence of the phage red gene that facilitates recombination between homologous sequences. We discovered that efficient recombination requires the presence of a λ phage gene, either *red* [Shulman et al., 1970] or *rap* (that maps to the nin5 region deleted in most λ cloning vectors) [Lutz et al., 1987; Kurnit and Seed, 1990]. In the case of Sumo15A, the presence of the λ phage *red* gene product ensures that recombination occurs efficiently on a theoretical basis; we verified this point experimentally as well [Kurachi et al., 1989]. Selection against phasmids that carry supF has proven more elusive, but has been realized finally with the construction of the Sumo15A phasmid vector [Kurachi et al., 1989] and the *E. coli* host, DK37 [Kurnit, 1989]. Paralleling the construction of Sumo15A, we synthesized an *E. coli* host to facilitate counterselection against supF. Derived from strain N953 [Russell et al., 1970], DK37 has the following genotype: $\Delta(\text{srl-recA})306$, *galKU42am*, *galE*⁻, *trp*^{-am}, *ara*^{-am}, *lacYT6R* (ϕ 80wt) [Russell et al., 1970; Willis et al., 1981; Kurachi et al., 1989; Kurnit, 1989]. Thus, the important features of DK37 are:

1. ϕ 80-derived λ phages (such as Sumo15A) cannot replicate lytically because of the expression of ϕ 80 immunity from the integrated ϕ 80 prophage.
2. Forward selection for supF is available from the presence of the *trp*^{-am} allele, whose suppression is selected for on plates that lack tryptophan.
3. Back-selection against supF is provided by "galactocide" counterselection: expression of galactokinase (afforded by suppression of the *galKU42am* gene by supF) is lethal in a galactose epimerase mutant (such as DK37) grown on plates containing galactose due to accumulation of toxic UDP-galactose. This counterselection against supF can be accomplished only for plasmids carrying supF. Therefore, we designed Sumo15A so that it could replicate as a plasmid in DK37 which expresses the ϕ 80 immunity gene product. Incorporation of the pSC101 replicon (as pGB2) into Sumo15A enables this

phasmid vector to propagate as a pSC101-driven phasmid replicon when the ϕ 80 gene product is present and turns off the λ -based phage replicon of Sumo15A.

4. DK37 is a deletion *recA*⁻ host, so that rearrangement of plasmids growing in this host are minimized [Kurnit, 1989].

Since the initial back-selection event involves the *recA*⁺-mediated excision of supF, we devised a rapid protocol to permit this event. We drop titer the Sumo15A phage carrying supF on a mixed lawn of the *recA*⁺ supF nonselective host LE392 and the *recA*⁻ selective host DK37. The blue phages from DK21 (that carry supF) are plated on this mixed lawn. Replication through the LE392 host occurs as a phage, allowing a small fraction of the phages to undergo *recA*-mediated recombination-based excision of the supF plasmid harboring an expressed genomic insert on chromosome 21. Replication of Sumo15A through the DK37 host occurs only as a plasmid that has lost the supF gene by recombination in LE392: supF is lethal to DK37 grown on galactose plates.

In this manner, we select first on DK21 for phasmids that incorporated supF by recombination, and then select on a mixed lawn of LE392 and DK37 for phasmids that lost supF by recombination (Fig. 2). By demonstrating that both forward and reverse reactions may be accomplished using this format, we documented that the strains we constructed may be used for recombination-based selection and counterselection of genic libraries by genomic inserts on chromosome 21.

Application to Chromosome 21

DNA from chromosome 21 has been isolated by flow sorting and cloned in cosmid vectors in other laboratories kind enough to provide us with these cosmids enriched in human chromosome 21 sequences. Our protocol is as follows: DNA is prepared from a given cosmid, cut with a restriction endonuclease, and probed with radiolabeled total human DNA, to detect repeated human DNA sequences [Britten and Kohne, 1968]. A non-repeated human DNA insert is thereby isolated, radiolabeled, and mapped on a panel of rodent-human somatic cell hybrids. We are particularly interested in sequences that map to two distinct parts of chromosome 21: the DS region of 21q22 and the pericentromeric region to which familial Alzheimer disease maps. Once a cosmid is identified that maps to a particular region of chromosome 21, we will use the recombination-based assay to ask whether any part of a particular sequence is represented in a cDNA library, i.e., is transcribed. We chose fetal brain and HeLa cells because of the complexity of these tissues, and because the brain represents the most significant organ affected by DS. The experiment will be to take a genomic fragment of nonrepeated chromosome 21 DNA [Neve and Kurnit, 1983], clone it in a plasmid with supF and chloramphenicol resistance, and ask by recombination if the genomic sequence is transcribed (i.e., whether it is present in a cDNA library). In this manner, we propose to identify sequences that are transcribed from specific regions of chromosome 21.

Alignment of Our Screen for Expressed Sequences With the Genomic Initiative on Chromosome 21

Significant energy and time will be expended to clone and sequence this smallest human autosome. Our methodology was designed to merge our interests, i.e., the isolation of expressed sequences on chromosome 21, with this genomic effort. This larger genomic effort should result in the sequencing of chromosome 21. However, neither the tissues nor the levels of transcription of potential genes, identified by sequencing, will be known. Our assay is designed to yield this information, in addition to the actual expressed sequences. In this manner, we propose to determine the level of transcription of chromosome 21, and to isolate the sequences encoded by this chromosome.

ACKNOWLEDGMENTS

Supported by March of Dimes grant 6-501. D.M.K. is an Investigator and G.D.S. is an Associate, Howard Hughes Medical Institute.

REFERENCES

- Blattner FR, Williams BG, Blechl AE, Denniston-Thompson K, Faber HE, Furlong LA, Grunwald DJ, Kiefer DO, Moore DD, Schumm JW, Sheldon EL, Smithies O (1977): Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. *Science* 196:161-169.
- Britten RJ, Kohne DE (1968): Repeated sequences in DNA. *Science* 161:529-540.
- Churchward G, Belin D, Nagamine Y (1984): A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors. *Gene* 31:165-171.
- Ikedo H, Aoki K, Naito A (1982): Illegitimate recombination mediated in vitro by DNA gyrase of *Escherichia coli*: Structure of recombinant DNA molecules. *Proc Natl Acad Sci USA* 79:3724-3728.
- Kurachi S, Baldori N, Kurnit DM (1989): Sumo15A: A lambda phasmid that permits manipulation of cloned inserts via selection and counterselection for supF. *Gene*, 85:35-43.
- Kurnit DM (1989): RecA deletion strains of *Escherichia coli* that are highly competent for transformation and for in vivo packaging. *Gene* 82:313-315.
- Kurnit DM, Seed B (1990): An improved system for screening bacteriophage libraries by homologous recombination in vivo. *Proc Natl Acad Sci USA*, 87:3166-3169.
- Lutz CT, Hollifield WC, Seed B, Davie JM, Huang HV (1987): An improved lambda phage vector designed for screening DNA libraries by recombination in vivo. *Proc Natl Acad Sci USA* 84:4379-4383.
- Marvo SL, King SR, Jaskunas SR (1983): Role of short regions of homology in intermolecular illegitimate recombination events. *Proc Natl Acad Sci USA* 80:2452-2456.
- Neve RL, Kurnit DM (1983): Comparison of sequence repetitiveness of human cDNA and genomic DNA using the miniplasmid vector piVX. *Gene* 23:355-367.
- Neve RL, Stewart GD, Newcomb P, Van Keuren ML, Patterson D, Drabkin HA, Kurnit DM (1986): Human chromosome 21-encoded cDNA clones. *Gene* 49:361-369.
- Russell RL, Abelson JN, Landy A, Geftler ML, Brenner S, Smith JD (1970): Duplicate genes for tyrosine transfer RNA in *Escherichia coli*. *J Mol Biol* 47:1-13.
- Seed B (1983): Purification of genomic sequences from bacteriophage libraries by recombination and selection in vivo. *Nucleic Acids Res* 11:2427-2445.
- Shulman MJ, Hallick LM, Echols HM, Signer ER (1970): Properties of recombination-deficient mutants of bacteriophage lambda. *J Mol Biol* 52:501-520.
- Willis DK, Uhlin BE, Amini KS, Clark AJ (1981): Physical mapping of the srl-recA region of *Escherichia coli*: Analysis of transposon Tn10 generated insertions and deletions. *Mol Gen Genet* 183:497-504.