The Genomic Region Encompassing the Nephropathic Cystinosis Gene (*CTNS*): Complete Sequencing of a 200-kb Segment and Discovery of a Novel Gene within the Common Cystinosis-Causing Deletion

Jeffrey W. Touchman,[1] Yair Anikster,[2] Nicole L. Dietrich,[1] Valerie V. Braden Maduro,[3] Geraldine McDowell,[2] Vorasuk Shotelersuk,[2] Gerard G. Bouffard,[1] Stephen M. Beckstrom-Sternberg,[1] William A. Gahl,[2] and Eric D. Green[1,3,4]

[1]*NIH Intramural Sequencing Center, National Institutes of Health, Gaithersburg, Maryland 20877;* [2]*Heritable Disorders Branch, National Institute for Child Health and Development and* [3]*Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892*

Nephropathic cystinosis is an autosomal recessive disorder caused by the defective transport of cystine out of lysosomes. Recently, the causative gene (*CTNS*) was identified and presumed to encode an integral membrane protein called cystinosin. Many of the disease-associated mutations in *CTNS* are deletions, including one >55 kb in size that represents the most common cystinosis allele encountered to date. In an effort to determine the precise genomic organization of *CTNS* and to gain sequence-based insight about the DNA within and flanking cystinosis-associated deletions, we mapped and sequenced the region of human chromosome 17p13 encompassing *CTNS*. Specifically, a bacterial artificial chromosome (BAC)-based physical map spanning *CTNS* was constructed by sequence-tagged site (STS)-content mapping. The resulting BAC contig provided the relative order of 43 STSs. Two overlapping BACs, which together contain all of the *CTNS* exons as well as extensive amounts of flanking DNA, were selected and subjected to shotgun sequencing. A total of 200,237 bp of contiguous, high-accuracy sequence was generated. Analysis of the resulting data revealed a number of interesting features about this genomic region, including the long-range organization of *CTNS*, insight about the breakpoints and intervening DNA associated with the common cystinosis-causing deletion, and structural information about five genes neighboring *CTNS* (human ortholog of rat vanilloid receptor subtype 1 gene, *CARKL*, *TIP-1*, *P2X5*, and *HUMINAE*). In particular, sequence analysis detected the presence of a novel gene (*CARKL*) residing within the most common cystinosis-causing deletion. This gene encodes a previously unknown protein that is predicted to function as a carbohydrate kinase. Interestingly, both *CTNS* and *CARKL* are absent in nearly half of all cystinosis patients (i.e., those homozygous for the common deletion).

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. AF168787 and AF163573.]

Nephropathic cystinosis is a rare autosomal recessive, lysosomal storage disease with an incidence estimated at 1 per 100,000–200,000 live births (see http://www.ncbi.nlm.nih.gov/omim; OMIM 219800). The classic disorder is characterized clinically by renal tubular Fanconi syndrome in the first year of life, growth retardation in childhood, renal glomerular failure at ~10 years of age, hypothyroidism, and a variety of other complications, including photophobia and corneal crystal formation (Gahl 1986; Gahl et al. 1995). After renal transplantation, cystine accumulation continues in nonrenal organs, frequently causing a distal vacuolar myopathy (Charnas et al. 1994), swallowing difficulty (Sonies et al. 1990), or retinal dysfunction (Kaiser-Kupfer et al. 1986), and occasionally causing diabetes mellitus (Fivush et al. 1987), pancreatic exocrine insufficiency (Fivush et al. 1988), or neurological deterioration (Ehrich et al. 1979; Fink et al. 1989). These complications arise because defective lysosomal transport of the disulfide cystine (Gahl et al. 1982a) causes this amino acid to accumulate within the lyso-

[4]Corresponding author.
E-MAIL egreen@nhgri.nih.gov; FAX (301) 402-4735.

10:165–173 ©2000 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/00 $5.00; www.genome.org

somes of many different cell types, which then triggers cystine crystal formation (Gahl et al. 1982b). The cystine transporter is the first of many lysosomal membrane carriers to be characterized biochemically (Thoene 1992), and cystinosis is the most common of a group of lysosomal transport disorders (Gahl et al. 1995).

The gene altered in patients with cystinosis (*CTNS*) was recently identified by a positional cloning strategy (Town et al. 1998). *CTNS* is a 12-exon gene that is transcribed into a ~2.6-kb mRNA. The encoded protein, named cystinosin, consists of a predicted 367 amino acids, appears to be an integral membrane protein, and most likely functions as a cystine transporter. A number of cystinosis-causing *CTNS* mutations have now been reported (Shotelersuk et al. 1998a; Town et al. 1998). The most prevalent mutation reported to date is a large (>55-kb) deletion, with 33%–44% of affected patients being homozygous for this deletion (Town et al. 1998; Anikster et al. 1999). In addition, at least 11 other smaller disease-causing deletions have been reported (Shotelersuk et al. 1998a; Forestier et al. 1999), suggesting that this genomic region may be prone to rearrangement.

We sought to establish the long-range organization of the segment of chromosome 17p13 harboring *CTNS* and to determine the sequence of this clinically important gene and its surrounding DNA. Here we report the assembly of a detailed bacterial artificial chromosome (BAC)-based physical map encompassing *CTNS*. In addition, two BAC clones spanning >200 kb were sequenced to high accuracy, providing insight into the molecular architecture of the *CTNS* gene and the genomic segment commonly deleted in cystinosis patients.

RESULTS

Physical Mapping

Our goal was to construct a high-resolution, long-range physical map of the region of chromosome 17p13 containing *CTNS*. Specifically, we sought to isolate the region in overlapping BAC clones (Shizuya et al. 1992; Birren et al. 1999) and to order a large set of sequence-tagged sites (STSs) across the interval. Although this genomic segment has been isolated in yeast artificial chromosomes (YACs) (McDowell et al. 1996; Stec et al. 1996; Peters et al. 1997), few markers were available for BAC isolation and mapping. Consequently, we generated new STSs across the region using several sources of DNA sequence, including known genes (e.g., *ASPA*) and genetic markers (e.g., D17S2167, D17S2054, D17S1828), a YAC spanning the interval [CEPH YAC 767F9 (McDowell et al. 1996; Peters et al. 1997)], and BAC insert ends. Available human BAC libraries were screened by PCR- and hybridization-based methods for the available STSs. Following STS-content analysis, nascent contigs were assembled, and clones residing at contig ends were selected and used to derive additional BAC insert-end sequences. New STSs were developed from the latter and used to screen the BAC libraries again. This scheme was repeated in an iterative fashion, eventually allowing assembly of the contig map depicted in Figure 1.

The resulting BAC-based STS-content map contains 95 clones and provides ordering information for 43 STSs. The contig is estimated to span >1 Mb based on previous YAC-based mapping of the interval (McDowell et al. 1996; Peters et al. 1997). The average redundancy of BACs per STS is ~14; such redundancy provides strong support for the indicated BAC overlaps and deduced STS order.

Genomic Sequencing

Two overlapping BACs (RG147P12 and RG87B10; see Fig. 1), which together contain the entire *CTNS* gene, were sequenced to an estimated accuracy of >99.99% by a shotgun sequencing strategy (Wilson and Mardis 1997). The clone inserts were found to be 68,220 and 138,720 bp in size, respectively, and to overlap by 6703 bp. Thus, a total of 200,237 bp of nonredundant sequence was generated (GenBank accession no. AF168787). Comparison of the sequence with a collection of known human repetitive elements revealed that this genomic region is relatively rich in repeats (constituting 42.6% of the total sequence), in particular short interspersed repetitive elements (SINEs). *Alu* repeats comprise nearly 30% of the sequence (Table 1).

Genomic Organization of the *CTNS* Gene

Comparison of the *CTNS* cDNA sequence and the generated genomic sequence allows the precise structure of the gene to be deduced, including details about intron/exon organization (Table 2; Fig. 2). The published *CTNS* cDNA sequence (GenBank accession no. AJ222967) is distributed across 24,816 bp of genomic DNA [positions 72,070–96,885 (GenBank accession no. AF168787)]. This cDNA sequence matches our established genomic sequence throughout, except for a silent A:G substitution at nucleotide position 843 in exon 8 (of the cDNA sequence), the presence of an additional T residue at position 2273 in the 3′-untranslated region (UTR), and a G:A substitution at position 2594 in the 3′ UTR. Furthermore, based on the genomic sequence, intron 1 is 276 bp in length, shorter than that described previously (Town et al. 1998).

Deletion Breakpoint Mapping

The breakpoints of the most common cystinosis-causing deletion were identified and sequenced in numerous cystinosis patients and reported previously (Anikster et al. 1999; Forestier et al. 1999). The avail-