Cytogenetic and
Genome Research

Human chromosome 3: integration of 60 *Not*I clones into a physical and gene map

G.E. Sulimova,[a] A.S. Kutsenko,[b,c] E.R. Rakhmanaliev,[a] I.G. Udina,[a]
A.A. Kompaniytsev,[a] A.I. Protopopov,[c] E.V. Moisjak,[a] E.A. Klimov,[a]
O.V. Muravenko,[b,c] A.V. Zelenin,[b] E.A. Braga,[c,d] V.I. Kashuba,[c]
E.R. Zabarovsky[b,c] and L.L. Kisselev[b]

[a] Vavilov Institute of General Genetics;
[b] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow (Russia);
[c] Microbiology and Tumor Biology Center and Center for Genomics and Bioinformatics, Karolinska Institute,
Stockholm (Sweden);
[d] Russian State Genetics Center, Moscow (Russia)

Abstract. Sequence tagged sites generated for 60 *Not*I clones (*Not*I-STSs) from human chromosome 3-specific *Not*I-jumping and *Not*I-linking libraries were physically located using PCR screening of a radiation hybrid (RH) GeneBridge4 panel. The *Not*I map of chromosome 3 was generated using these RH-mapping data and those obtained earlier by FISH and sequencing of the corresponding *Not*I clones. The sequences of the *Not*I clones showed significant homologies with known genes and/or ESTs for 58 *Not*I-STSs (97%). These 58 *Not*I clones displayed 91–100% identity to 54 genes and 23 cDNA/EST clones. One known and two hypothetical protein-coding genes were localized for the first time and nine cDNA clones (unknown genes) were also carefully mapped only in this work. Three newly mapped genes are histone gene H1X (NR1-BK20C) and genes for hypothetical proteins THC1032178 and THC1024604 (NL1-243).

Copyright © 2002 S. Karger AG, Basel

During the past decade, human genomics has focused on various types of mapping, including gene mapping along chromosomes, mapping of unique sequences relative to each other, and mapping of expressed regions of the whole genome. As a result of these efforts, combined with large-scale random sequencing, a huge corpus of structural information has emerged being in many cases duplicated, overlapping, or even contradictory. The huge recent efforts deployed on sequencing the whole human genome (Lander et al., 2001; Venter et al., 2001) are extremely important, however, they have not ruled out the necessity of generating verified integrated maps. Short and long repeats of various types spread throughout the whole genome are sources of numerous errors that are difficult to identify with a shotgun sequencing strategy, but which become evident if mapping information is combined with the sequence information. Furthermore, problems caused by the existence of gene families widespread among the higher vertebrates, as well as pseudogenes, chimeric genes, and others, are much easier to resolve using integrated gene maps. We have to keep in mind that even absolutely correct and long nucleotide sequences may be incorrectly mapped along the chromosomal DNA if the appropriate accompanying mapping information is absent or ignored. Gene mapping would be accelerated by the existence of universal gene markers to distinguish coding from non-coding regions of the genome. A good example of such markers is CpG islands (Bird, 1987) that are tightly associated with genes. It was directly demonstrated that *Not*I restriction sites are associated with CpG islands (Larsen et al., 1992; Allikmets et al., 1994; Protopopov et al., 1996). This means that unique

KARGER

Fax + 41 61 306 12 34
E-mail karger@karger.ch
www.karger.com

© 2002 S. Karger AG, Basel
0301-0171/02/0983-0177$18.50/0

Accessible online at:
www.karger.com/cgr

sequences flanking *Not*I restriction sites (*Not*I-STSs) could serve as universal markers for the major portion of human genes and probably for mammalian genes in general.

Recently, we generated a set of 22,551 (16.2 Mbp) unique human *Not*I flanking sequences (Kutsenko et al., 2002). More than 40% of the sequences had regions with significant similarity to known proteins and expressed sequences. The data demonstrated that regions flanking *Not*I sites are less likely to form nucleosomes efficiently and resemble promoter regions. We experimentally tested more than 200 *Not*I linking clones and all of them contained genes (see for example Allikmets et al., 1994; Protopopov et al., 1996; Kashuba et al., 1999, 2002). The draft human genome sequence contained 55.7% of the *Not*I flanking sequences and Celera's database contained matches to 57.2% of the clones (Kutsenko et al., 2002). The data suggested that the shotgun sequencing approach used to generate the draft human genome sequence resulted in a bias against cloning and sequencing of *Not*I flanks. A rough estimation (based primarily on chromosome 21 and 22) is that the human genome contains 15,000–20,000 *Not*I sites, of which 6,000–9,000 are unmethylated in any particular cell. These flanking sequences were already helpful in the isolation and mapping of new genes and resolving ambiguities in chromosome 3 maps (see Kutsenko et al., 2002; Kashuba et al., 1999, 2002; Protopopov et al., 1996, 2003). *Not*I clones can serve as STSs that can be mapped precisely using PFGE and FISH. We think that the *Not*I clones will also be helpful as probes to close existing gaps in the draft human genome sequence and in estimating the completeness of the human genome sequence due to the independent approach used in this study.

Methylation, deletions, and amplifications of cancer genes constitute important mechanisms in carcinogenesis. We have shown that *Not*I clone microarrays can simultaneously detect copy number changes and methylation. *Not*I microarrays offer a powerful tool with which to study carcinogenesis (Li et al., 2002) and therefore location of *Not*I and associated genes will be important not only for structural but also for functional genomics.

Until now, information regarding the location of *Not*I sites on human chromosome 3 was mainly derived from fluorescence *in situ* hybridization (FISH) studies (see for example, Protopopov et al., 1996; Fedorova et al., 1997) and from yeast artificial chromosome (YAC)-contigs (Gemmil et al., 1995). The sensitivity of FISH analysis is insufficient for precise physical and gene mapping, and numerous rearrangements severely disturb the genuine *Not*I site patterns in mega-YACs.

The objective of this work was to start creating an integrated gene map of human chromosome 3 using *Not*I clones, radiation hybrid (RH) mapping and the data on FISH and sequencing of these clones obtained earlier (Kashuba et al., 1995, 1999; Protopopov et al., 1996; Wei et al., 1996; Fedorova et al., 1997). Thus in this work 60 *Not*I clones (four of which were absent in assembled chromosome 3 sequence in January 2003) and 54 genes and 23 cDNAs were incorporated into the integrated map.

Materials and methods

Analysis of NotI-linking clones and sequencing

Chromosome 3-specific *Not*I-jumping and -linking libraries were constructed using MCH903.1 and MCH924.4 cell lines, as described previously (Zabarovsky et al., 1994a; 1996). Both MCH903.1 and MCH924.4 are mouse–human microcell hybrids derived from the HFDC and HHW1108 normal human diploid fibroblast cell lines, respectively, and contain, as their only human component, a single copy of human chromosome 3 (Zabarovsky et al., 1993; Wang et al., 1994). Chromosome 3-specific clones were selected from the plasmid libraries by differential hybridization to ^{32}P-labeled total human and mouse DNA. The selected clones hybridized to human DNA, but not to murine DNA (Zabarovsky et al., 1993; Wang et al., 1994). The nomenclature of the clones has been described previously (Kashuba et al., 1999). DNA isolation and hybridization were performed according to standard procedures (Sambrook et al., 1989). Sequencing was carried out as described earlier (Zabarovsky et al., 1994b; Kashuba et al., 1999). Each clone was sequenced from both ends. The sequencing reactions were resolved on an ABI377 Sequencer (Perkin Elmer, Norwalk, CT, USA) according to the manufacturers' protocols.

Sequence analysis

The analysis of sequences was performed at the Karolinska Institute Bioinformatics Facility (http://kisac.cgb.ki.se), using local versions of programs and public databases. Nucleotide similarity searches were performed with the BLAST 2.0 program (Altschul et al., 1990; 1997). Initially, masking the repeats was performed by the RepeatMasker program (Smit and Green, at http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker). Only interspersed repeats were masked, and low complexity DNA and simple repeats (tandem repeats) were not masked. Then a comparison was made between the initial *Not*I clone sequences and both the unfinished and finished human genomic sequences. For subsequent analyses, extended sequences with homology to the *Not*I clones were taken from both the finished and the unfinished human genome. Each new sequence included *Not*I sites with the surrounding 1,000 bp in both the 5′ and 3′ directions. If there were adjacent *Not*I sites with a distance between them of less than 2,000 bp, both of them were considered as one sequence. In that way, we generated the set of sequences surrounding *Not*I sites (*Not*I genomic sequences) that were equivalent to the set of initial clones, but these *Not*I genomic sequences were more precise and extended.

A new BLAST homology search was performed for the masked *Not*I genomic sequences against the EMBL and EST databases. Extra homology searches were carried out in the SRS database, to identify the known markers. When the known markers for the *Not*I genomic sequences were not found, we lengthened the *Not*I genomic sequences to encompass the total human genomic contigs containing these *Not*I sites.

The work was done using EMBL sequence databases, release 73 (December 2002). Mapping *Not*I sequences was performed with publicly available human genome mapping services (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi, MapView build 31, January 2003)

RH-mapping

To localize *Not*I-STSs, PCR-based screening of a somatic cell radiation hybrid (human/hamster) GeneBridge4 RH-panel (Research Genetics Inc., USA) was applied as described earlier (Sulimova et al., 1999).

PCR was performed in 12.5-µl reactions, using a PTC-100™ thermocycler (MJ Research Inc., USA). Some modifications were introduced, including the "hot start" technique, as recommended in the instructions for Maxi-Taq™, manufactured by Biokom Inc. (Russia). The amount of DNA per tube was 8–10 ng. The nucleotide sequences of the primers used in these studies are available in Table 1 at http://www.mtc.ki.se/groups/zabarovsky/. Conditions for annealing, priming, and other parameters were optimized using total human and hamster DNAs. The analysis of PCR products using agarose and polyacrylamide gels was performed following standard protocols (Sambrook et al., 1989).

The results of PCR screening of the RH-panel were analyzed with RHMAPPER through the server of the Whitehead Institute/MIT Center for Genome Research (http://www-genome.wi.mit.edu). To reconcile these data with previously reported information regarding maps of the chromosome, three additional sources were used (http://www.ncbi.nlm.nih.gov/; http://www.stanford.edu and http://www.tigr.org/).

Results and discussion

Generation of NotI-STSs and their analysis

NotI-linking and NotI-jumping libraries specific for human chromosome 3 were sequenced around NotI sites (see Kashuba et al., 1995, 1999; http://www.mtc.ki.se/groups/zabarovsky/). These sequences were used to generate 70 NotI-STSs for 60 individual NotI clones (see Table 1 at http://www.mtc.ki.se/groups/zabarovsky/).

The sequences of the NotI clones were compared with known genes and ESTs available in public databases. Homology searches were performed as described in "Materials and methods". Only one homologous sequence with the highest degree of identity to the individual NotI clones has been listed in Table 1 (see Table 2 at http://www.mtc.ki.se/groups/zabarovsky/ for more details). Sequences with levels of identity below 90% were discarded. Significant homologies were found with genes and ESTs for 58 NotI clones (97% of the total number of clones). Homologies with NotI clones were identified for 54 genes (48 of which had 93–100% identity in sequences from 91 to 1144 nucleotides in length) and 23 cDNA/EST clones (91–100% identity in sequences from 156 to 541 nucleotides in length). Six NotI clones (J32-135H, NL1-201, NL1-232, NL3-008, NL3-009, and NLJ-003) exhibited short (55–83 bp) but 100% identical sequences with known genes.

As became evident from the analysis (see Table 2 at http://www.mtc.ki.se/groups/zabarovsky/), flanking sequences of NotI sites could be homologous to two ESTs, or to a gene and an EST, or to two genes (AP-19, NL1-243, NL2-007, NL3-004, NRL-097 and NRL1-1). In five cases, two NotI-STSs were found to be homologous to two different regions of a single gene: AP-19 and NL1-201 were homologous to ABCC5, J32-180 and NL1-358 to SCA7, NRL-097, NL1-216 to ARP and NL1-401 and NRL-402 to ITGA9 and NL1-155 and NL2-316 to SSR3.

Comparisons of the new (Table 1) and previous (Kashuba et al., 1999) data revealed novel homologies between 26 NotI clones and genes/cDNAs, previously not identified. Among these homologies 16 were completely new and in ten cases NotI clones were linked to genes although earlier identity was found only to unknown cDNA clones.

Two NotI-STSs (NL1-256 and NLM-067) showed no significant homology with genes or ESTs present in the databases. Interestingly, the number of clones exhibiting gene-marker properties increased from 68% in August 1999 to 97% in January 2003, due to the enrichment of available databases. It is easy to foresee that in the near future, nearly all NotI-STSs will be considered genuine gene markers, and their chromosome locations will map the given gene.

Two NotI-STSs (AP-20 and NL1-256) showed no homology with human genome sequences in the draft human genome

Table 1. Ordering of NotI-STSs on human chromosome 3

No	Name[a]	Homolog (gene or EST)[a]	Distance (in cR) from framework RH-markers	Position in cR[b] WI-RH	GM99'-GB4 according to RH-data or (Contig No[c])	Reference interval (in cM)	Cytogenetic location according to: RH-data[c]	FISH[d]	Gene location[e]	Genome Position (according to Blast) Genomic Contig	Position, Mbp
1	NL1-229	LOC131470	5.76 from AFMA216zg1 8.56 from CHLC.GATA51A05	54.2 47.5	45.9 40.3	D3S3591 - D3S1283 (19.0 - 46.8)	3p26.1	3p25	3p25.3	NT_005927	10.7
2	NL1-106	TADA3L	-	53.7-55.4* 55.1*	46.81* 46.81*	D3S3591 - D3S1283 (19.0 - 46.8)	3p26*	3p25.2-p25.1	3p25.3	NT_005927	11.1
3	J32-159L[f]	cDNA clone 5' (AA234690)	-	53.7-55.4*	46.81*	D3S3591 - D3S1283 (19.0 - 46.8)	3p26*	3p21.32-p21.2	no data	NT_005927	11.0-11.6
4	NL1-177	LOC63110	-	73.1*	67.51 – 68.05*	D3S3591 - D3S1283 (19.0 - 46.8)	3p25.3*	3p25- p24.2	3p24.3	NT_005927	18.5
5	AP-1	cDNA clone 3' (AA626023)	3.15 from WI-7376 3.87 from WI-4218	73.2 72.4	68.3	D3S3591 - D3S1283 (19.0 - 46.8)	3p25.3	3p25	no data	NT_005927	18.6
6	NL1-024	cDNA clones 3' (AA405005, AA700946)	-	-	124.4 (Contig 1)	D3S1609 – D3S1260 (49.5 – 57.8)	3p22.1	3p22-21.33**	no data	NT_034532	31.4
7	AP-40	LOC166003	-	-	125.5 (Contig 1)	D3S1609 – D3S1260 (49.5 – 57.8)	3p22.1	3p22-21.33**	3p21.1	NT_005580	38.1
8	NL1-401	ITGA9	-	-	127.1 (Contig 1)	D3S1609 – D3S1260 (49.5 – 57.8)	3p21.3	3p22-21.33**	3p21.3	NT_005580	38.5
9	NRL-402	ITGA9	-	-	127.1 (Contig 1)	D3S1609 – D3S1260 (49.5 – 57.8)	3p21.3	3p22-p21.33	3p21.3	NT_005580	38.9
10	NLJ-003	HYA22	-	-	128.3 (Contig 1) 133.68*	D3S1260 – D3S3582 (57.8 – 65.1)	3p21.3 3p21.3*	3p22-21.33**	3p21.3	NT_005580	39.0
11	NL3-019	DLEC1	-	-	129.6 (Contig 1)	D3S1260 – D3S3582 (57.8 – 65.1)	3p21.3	3p22-p21.33	3p22-p21.3	NT_005580	39.1-39.2
12	AP-20	SCN5A	2.12 from D3S1561 4.92 from WI-4073	140.9 138.6 147.1*	131.0 128.9 (Contig 1) 133.5*	D3S1260 – D3S3582 (57.8 – 65.1) D3S1260 – D3S3582 (57.8 – 65.1) *	3p21.33 3p21.3*	3p21.33**	3p21	no data	no data
13	NL3-003	GORASP1	2.22 from D3S1561 5.02 from WI-4073	141.0 138.7	131.1 129.0 (Contig 1) 133.5 – 135.4*	D3S1260 – D3S3582 (57.8 – 65.1)	3p21.33 3p21.3*	3p21.33**	3p22-p21.33	no data (AC094020 RP11-425J9)	no data

➤

Table 1 (continued)

No	Name[a]	Homolog (gene or EST)[a]	Distance (in cR) from framework RH-markers	Position in cR[b] WI-RH	GM99'-GB4 according to RH-data or (Contig No[c])	Reference interval (in cM)	Cytogenetic location according to: RH-data[c]	FISH[d]	Gene location[e]	Genome Position (according to Blast) Genomic Contig	Position, Mbp
14	NL1-308	MOBP	-	-	133.5 (Contig 1) 133.09*	D3S1260 – D3S3582 (57.8 – 65.1)	3p21.33 3p21.3*	3p21.33**	3p21.33	NT_005498	39.8
15	NR1-NK20	GC20	-	-	137.45 (Contig 1)	D3S1260 – D3S3582 (57.8 – 65.1)	3p21.32	3p21.3	3p21.33	NT_006036	40.8
16	NL1-256	-	5.23 from D3S3564	166.7	146.2	D3S1260 – D3S3582 (57.8 – 65.1)	3p21.32	3p22-21.33**	no data	no data	no data
17	NL3-007	Similar to LOC51054	2.53 from AFMB362wb9	182.3	148.2 (Contig 2)	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31	3p21.3**	12q	NT_009781 (chr. 12)	45.5
18	NL3-006	KIAA0540	-	-	148.2 (Contig 2)	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31	3p21.3**	3p21.31	NT_029949	47.2
19	NLM-067	-	2.53 from AFMB362wb9	182.3	148.2 (Contig 2)	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31	3p21.3**	no data	NT_029949	47.3
20	NL1-232	CGI-58	6.08 from D3S3564	167.6	147.0 148.2 (Contig 3) 145.01*	D3S3582 – D3S1588 (65.1 – 67.7) D3S1260 – D3S3582 (57.8 – 65.1)	3p21.31 3p21.3*	3p21.3**	3p25.3-p24.3	NT_005825	45.6
21	NL3-009	KIAA0096	-	- 163.6*	148.2 (Contig 3) 142.6-143.9*	D3S3582 – D3S1588 (65.1 – 67.7) D3S1260 – D3S3582 (57.8 – 65.1)	3p21.31 3p21.31*	3p21.3**	3p24.3-p22.1	NT_005825	43.9
22	NL2-007	KIAA0116, ZDHHC3	-	173.5*	145.29* 149.4-149.9*	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31*	3p21.3**	3p21.32-p21.31 3p21.32	NT_005825	44.3
23	NL1-320	LIMD1	-		149.14*	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31*	3p21.3	3p21.3	NT_005825	44.8
24	J32-135L[f]	ARHA	3.87 from AFMB362wb9	182.7	148.5	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31	3p21.3**	3p21.3	NT_022439	49.3
25	J32-135H[f]	LOC160788	3.87 from AFMB362wb9 2.94 from WI-9324	182.7 174.1	148.5 145.6	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31	3p21.3**	12p11	NT_022439	50.1
26	J32-032H[f]	E2IG3	4.60 from AFMB362wb9	184.4	149.9	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.31	3p21	3p21.31	NT_005986	52.7
27	NL3-005	Similar to ATRN	-	-	159.0 (Contig 4)	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.2	3p21.32-p21.2**	no data	NT_005832	142.8
28	NRL1-1	RBM5, SEMA3F	-	-	160.0 (Contig 4) 160.35*	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.2 3p21.2*	3p21.32-p21.2**	3p21.3 3p21.3	NT_006014	50.7
29	NL3-001	GNAI2	-	- 204.9*	160.3 (Contig 4) 160.35*	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.2 3p21.2*	3p21.32-p21.2**	3p21	NT_006014	50.6
30	NL1-210	MAPKAPK3	-	- 207.7*	161.0 (Contig 4) 160.24* 162.5-162.7*	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.2 3p21.2*	3p21.32-p21.2**	3p21.3	NT_006014	50.2
31	NL1-216	ARMET	3.36 from WI-9590 10.54 from HLC.GATA87B02	213.0 210.6	164.3 162.5 (Contig 5) 160.3-161.1*	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.1 3p21.2*	3p21.32-p21.1**	3p21.1	NT_005986	51.0
32	NRL-097	ARMET	-	-	163.0 (Contig 5)	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.1	3p21.32-p21.1**	3p21.1	NT_005986	51.0
33	NRL-143	SPC12	-	-	163.5 (Contig 5)	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.1	3p21.2-p21.1**	3p21.31	NT_005986	52.7
34	NL1-245	ACY1	-	-	164.0 (Contig 5) 160.02*	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.1 3p21.1*	3p21.2-p21.1**	3p21.1	NT_005986	53.4
35	NL1-243	Similar to THC1024606, Similar to THC1032178	2.33 from WI-3521 3.67 from WI-9590	213.9 213.3	164.9 164.5 (Contig 5)	D3S3582 – D3S1588 (65.1 – 67.7)	3p21.1	3p21.1-p14.3**	no data	NT_005986	53.4
36	NL3-008	DOCK3	1.61 from GCT4B10 10.43 from CHLC.GATA87B02	222.7 245.7	171.7 189.4 (Contig 6) 160.98*	D3S1289 – D3S1547 (69.1 – 77.4) D3S3582 – D3S1588 (65.1 – 67.7)*	3p21.1 3p21.1*	3p21.1-p14.3**	3p21.31	NT_005986	53.9
37	NL3-004	Similar to KIAA1157	10.43 from CHLC.GATA87B02	245.7	189.4 (Contig 6)	D3S1289 – D3S1547 (69.1 – 77.4)	3p21.1	3p21.1-p14.3**	12q13.13	NT_005986	54.0

Table 1 (continued)

No	Name[a]	Homolog (gene or EST)[a]	Distance (in cR) from framework RH-markers	WI-RH	GM99'-GB4 according to RH-data or (Contig No[c])	Reference interval (in cM) RH-data	[c]	FISH[d]	Gene location[e]	Genomic Contig	Position, Mbp
38	NL2-008	cDNA clone 3' (AA013220)	10.54 from HLC.GATA87B02	245.8	189.5 (Contig 6)	D3S1289 - D3S1547 (69.1 – 77.4)	3p21.1	3p21.1-p14.3**	no data	NT_005986	54.1
39	AP-32	cDNA clone 3' (AI217472) 5' (BI461949)	10.43 from CHLC.GATA87B02	245.7	189.4 189.8 (Contig 6)	D3S1289 - D3S1547 (69.1 – 77.4)	3p21.1	3p21.1-p14.3**	no data	NT_005986	54.2
40	**AP-2**	AR F4	0.80 from WI-3771 2.22 from GCT4B10	225.1 223.3	173.6 172.2 191.54*	D3S1588 - D3S1289 (67.7 - 69.1) D3S1289 – D3S1547* (69.1 – 77.4)	3p21.1 3p21.1*	3p21.1-p14.3**	3p21.2-p21.1	NT_005787	56.7
41	**NLM-223**	**Similar to LOC57406**	-	-	190.17*	D3S1289 – D3S1547 (69.1 – 77.4)	3p21.1*	3p21.1-p14.3**	3p21.1	NT_005787	57.2
42	J32-184R[f]	cDNA clone 3' (AI204459)	-	-	191.5*	D3S1289 – D3S1547* (69.1 – 77.4)	3p21.1*	no data	no data	NT_005999	64.6
43	J32-180H[f]	**SCA7**	-	-	190.9-198.0*	D3S1547 - D3S1261 (77.4 – 97.2)	3p21.1-p14.3*	3p21.1-p14.3**	3p21.1-p12	NT_005999	64.7
44	**NL1-358**	**SCA7**	-	-	191.3-198.0*	D3S1547 - D3S1261 (77.4 – 97.2)	3p21.1-p14.3*	3p14** 3	p21.1-p12	NT_005999	65.0
45	**NR1-BK20**	**H1x**	14.16 from CHLC.GATA68D03	457.1	384.0	D3S1291 - D3S1302 (121.9 – 125.5)	3q13.1-q13.2	no data	no data	no data AC137695 (RP13-685P2)	no data
46	NLM-187	cDNA clone 5' (AA301132)	3.98 from WI-9663 5.23 from WI-1780	497.4	405.6	D3S1302 - D3S1610 (125.5 – 130.1)	3q13.1-q13.2	3q13.2-q13.3	no data	NT_005795	115.6
47	**NL1-155**	**SSR3**	5.23 from WI-3188 7.26 from WI-6382	520.1	417.4	D3S1278 - D3S1303 (131.8 – 138.4)	3q13.2-q13.3	no data	no data	NT_005612	157.2
48	924-059	cDNA clone 3' (BE676755)	-	-	439.06*	D3S1267 - D3S1269 (141.1 – 142.2)	3q13*	3q28-q29	no data	NT_005543	130.7
49	**NLM-007**	**RAB7**	- -	-	447.61* 458.3-458.7*	D3S1269 - D3S3606 (142.2 – 146.0) D3S3606 - D3S3554 (146.0 – 156.2)	3q21* 3q21.3*	3q21 3	q22.1	NT_005612	134.6
50	**NL2-316**	**SSR3**	4.19 from D3S1570	710.9	577.2 577.3-577.4* 574.92*	D3S1275 - D3S1605 (176.1-177.4) D3S1280 – D3S1275* (174.3-176.1)	3q25.2 3q25.2*	3q23-q24	3 q26.1	NT_005612	157.2
51	**NRL-090**	**GMPS**	5.66 from WI-3847	775.0	635.6	D3S1553 – D3S1580 (178.1 – 213.7)	3q26.31	3q24**	3q24	NT_005612	166.1
52	**NLM-246**	**Similar to HSA9947**	6.29 from WI-6365 6.51 from WI-6165	853.3 835.0	687.9 686.1	D3S1553 – D3S1580 (178.1 – 213.7)	3q27.2	3q28-q29	1p36	NT_022406	188.3
53	924-066	**LAMP3**	-	-	682.99*	D3S1553 – D3S1580 (178.1 – 213.7)	3q27.2*	no data	3q26.3-q27	NT_022676	189.3
54	**NL1-201**	**ABCC5**	0.0 from D3S1571 12.67 from WI-6365	847 867.4	691.7 692.7 689.63*	D3S1553 – D3S1580 (178.1 – 213.7)	3q27.2-q28 3q27*	3q27-q28**	3q27	NT_022676	189.3
55	**AP-19**	**ABCC5, ETV5**	6.29 from WI-6365 5.87 from WI-6165	853.3 834.4	687.9 685.6 685.1-689.6* 697.3*	D3S1553 – D3S1580 (178.1 – 213.7)	3q27.2-q28 3q27.2-q28*	3q28 3	q27 3q28	NT_005911	192.0
56	924-069	**CLCN2**	-	-	695.2-698.7*	D3S1553 – D3S1580 (178.1 – 213.7)	3q28*	no data	3q27-q28	NT_029253	190.5
57	AP-34	cDNA clone 3' (AI827719)	4.60 from WI-1189 13.58 from WI-9695	871.1 874.5	717.4 719.6	D3S1314 - D3S1601 (218.3 - 220.4)	3q29	3 q28-q29	-	NT_005535	203.1
58	**NLM-084**	**Similar to FOXD4**	2.22 from D3S1314 4.92 from WI-1189	874.9 871.4	719.9 717.8	D3S1314 - D3S1601 (218.3 - 220.4)	3q29	3q29	9p11-q11	no data	no data
59	**NL2-252**	**TFRC**	11.99 from WI-9695	876.1	73 4.2 726.38*	D3S1265 –chr3-qTEL (228.0 - 233.0)	3q29 3q29*	3q29** 3	q26.2-qter	NT_033013	201.4
60	**NLM-216**	**FLJ20522**	-	883.7-905.1*	732.68*	D3S126 5 –chr3-qTEL (228.0 - 233.0)	3q29*	3p21.33-p23	3q29	NT_022928	201.7

[a] The names and accession numbers of NotI-STSs **homologous to genes** are given in bold and homologous to ESTs are underlined,

[b] Position in cR and cytogenetic location according to RH-data obtained from public databases are designated by *.

[c] NotI clones belonging to the same contig of NotI jumping/linking clones are indicated by numbers from 1 to 6 according to Kashuba et al. (1999)

[d] Cytogenetic location and ordering of NotI-STSs obtained earlier by us; clones mapped using di- and tricolour FISH Kashuba et al. (1999) are designed by **.

[e] According to literature data.

[f] J32 are NotI jumping clones. J32-159L, J32-032H, J32-184R clones represent chimeric NotI jumping clones. Only half of the insert (BamHI-NotI fragment shown in the Table, see Zabarovsky et al., 1996) is located on chromosome 3.
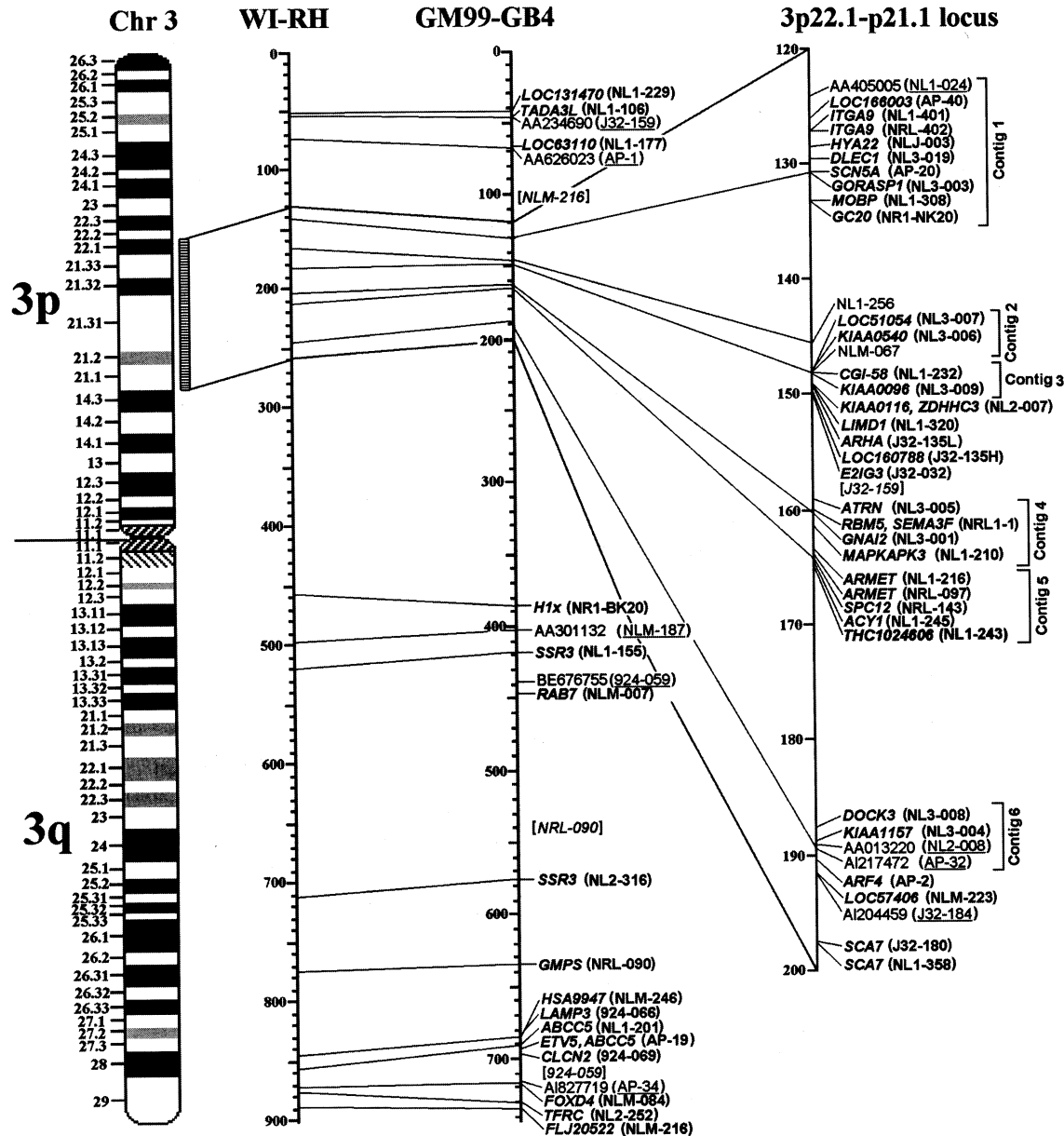
Fig. 1. Integrated *Not*I map of human chromosome 3. From left to right: ideograms of G-banding, RH-map of Whitehead Institute (WI-RH), GeneMap'99 (GM99-GB4) and 3p22.1 → p21.1 locus. The distances are expressed in centirays, cR. *Not*I-STSs homologous to genes are bold and homologous to expressed sequences (ESTs) are underlined. For J32-159, NLM-216, 924-059 and NRL-090 markers two positions are indicated. The second positions (putative pseudogenes) are given in brackets without exact mapping of the markers on the RH-map, since their locations are based solely on the two-colour FISH data.

sequence (January, 2003). Sequences from two other *Not*I linking clones (NL3-003 and NL1-BK20) are present in only non-assembled and non-mapped sequences. NL1-256 *Not*I clone showed no significant homology with any sequence represented in public databases. These data demonstrate that some gene fragments and anonymous ESTs submitted to the public databases are absent in the draft human genome sequence.

Mapping NotI-STSs and construction of an integrated map of human chromosome 3

Integration of *Not*I-STS into an RH map was done using PCR screening of a GeneBridge4 panel. For more convenient comparison of our new data with that collected earlier in the databases, we converted the Whitehead Institute RH-map (WI-RH-Map) into the GeneMap99-GB4 (GM99-GB4) (Table 1). This rescaling should not generate any essential errors because the physical locations for the majority of framework markers are known for both RH-maps.

The physical mapping of contigs of *Not*I jumping/linking clones was achieved by means of one or two *Not*I-STSs, which belonged to the given contigs. These contigs represented overlapping *Not*I jumping and linking clones. The locations of other *Not*I-STSs of the same contig were calculated, taking into account the size of the contig and the distances between the clones forming the contig that were established by PFGE (Kashuba et al., 1995, 1999; Wei et al., 1996).

When the physical locations of *Not*I-STSs mapped by us, and the corresponding genes were already known, we noted in most cases a good coincidence between the position of the gene and its marker. Therefore, the known location of certain genes on the RH-map helps to locate the corresponding *Not*I-STS markers, which are highly homologous to these genes in the same positions, and to introduce these data into Table 1 (NL1-106, NL1-177, NL1-320, NL2-007, NLM-223, and others). Furthermore, RH-mapping data combined with contig mapping of the 3p22→p21 locus (Kashuba et al., 1999) helped to refine the positions of markers located on the 3p22→p21 contigs and other markers (for example NL1-024, NL1-401, NLJ-003, NRL1-1, NL1-210).

In six cases we observed deviations in the locations of the genes/cDNAs and *Not*I markers by FISH and RH mapping. However, in one of these cases (NL1-232) our previous FISH mapping coincided with RH mapping. In another case (NRL-090) both our previous and literature FISH data were identical but differed from RH mapping. In four other cases (J32-159, NL1-155, NL3-005 and 924-059) FISH data were available only from one source: either from our previous FISH mapping of *Not*I clones or literature data.

Five other *Not*I-STSs (NL3-007, J32-135H, NL3-004, NLM-246 and NLM-084) have high homology with genes earlier mapped to other chromosomes (Table 1). Interestingly three of them were mapped to chromosome 12 (NL3-007, J32-135H and NL3-004). These results suggest that homologs (or pseudogenes) of these genes are located on human chromosome 3. Indeed, three of these *Not*I-STSs found identical sequences in chromosome 3 sequence (J32-135H, NL3-004 and NLM-246) and NLM-084 matched sequences from several human chromosomes.

These non-coincidences could be accounted for by various factors, such as the large size of some genes, the occurrence of related genes from the same gene family, pseudogenes, or gene duplications. For example, gene UNC93B1 originally was isolated using *Not*I clone NL1-304 (3p13→p12). Later, it was found that four highly homologous genes (93%–100% identity) located on four different chromosomes exist in the human genome (Kashuba et al., 2002).

One known and two hypothetical protein coding genes were localized for the first time and nine cDNA clones (unknown genes) were also carefully mapped only in this work. The three newly mapped genes are histone gene H1X (NR1-BK20C) and genes for hypothetical proteins THC1032178 and THC1024604 (NL1-243). As mentioned above, four *Not*I clones that were mapped here did not have matches to the assembled draft chromosome 3 sequence at all.

Both FISH and RH data on *Not*I-clones are summarized in Table 1 together with the D-loci intervals (in centimorgans) and sequencing data for human chromosome 3, to which the appropriate *Not*I-STSs have been mapped. As a consequence of the comprehensive experimental data the first integrated *Not*I/gene map of human chromosome 3 has been composed (Fig. 1).

References

Allikmets RL, Kashuba VI, Petterson B, et al: *Not*I linking clones as a tool for joining physical and genetic mapping of the human genome. Genomics 19:303–309 (1994).

Altschul SF, Gish W, Miller W, Myers EV, Lipman DJ: Basic local alignment search tool. J molec Biol 215:403–410 (1990).

Altschul SF, Madden TL, Schaffer AA, et al.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 25:3389–3402 (1997).

Bird AP: CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347 (1987)

Fedorova L, Kost-Alimova M, Gizatullin RZ, et al: Assignment and ordering of twenty new unique *Not*I linking clones containing cDNAs and human guanosine 5′-monophosphate synthase gene to human chromosome 3. Eur J hum Genet 5:110–116 (1997).

Gemmill RM, Chumakov I, Scott P, et al: A second-generation YAC contig map of human chromosome 3. Nature 377:299–319 (1995).

Kashuba VI, Szeles A, Allikmets R, et al: A group of *Not*I jumping and linking clones cover 2.5 Mb in the 3p21→p22 region suspected to contain a tumor suppressor gene. Cancer Genet Cytogenet 81:144–150 (1995).

Kashuba VI, Gizatullin RZ, Protopopov AI, et al: Analysis of *Not*I linking clones isolated from human chromosome 3 specific libraries. Gene 239:259–271 (1999).

Kashuba VI, Protopopov AI, Kvasha SM, et al: hUNC93B1: a novel human gene representing a new gene family and encoding an unc-93-like protein. Gene 283:209–217 (2002).

Kutsenko AS, Gizatullin R, Al-Amin AN, et al: *Not*I flanking sequences: a tool for gene discovery and verification of the human genome. Nucl Acids Res 30:3163–3170 (2002).

Lander ES, Linton LM, Birren B, et al: Initial sequencing and analysis of the human genome. Nature 409:860–921 (2001).

Larsen F, Gundersen G, Lopez R, Prydz H: CpG islands as gene markers in the human genome. Genomics 13:1095–1107 (1992).

Li J, Protopopov A, Wang F, et al: *Not*I subtraction and *Not*I-specific microarrays to detect copy number and methylation changes in whole genomes. Proc natl Acad Sci, USA 99:10724–10729 (2002).

Protopopov AI, Gizatullin RZ, Vorobieva NV, et al: High resolution FISH mapping of 50 *Not*I linking clones homologous to genes and cDNAs on human chromosome 3. Chrom Res 4:443–447 (1996).

Protopopov A, Kashuba VI, Zabarovska VI, et al: An integrated physical and gene map of the 3.5-Mb chromosome 3p21.3 (AP20) region implicated in major human epithelial malignancies. Cancer Res 63:404–412 (2003).

Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

Sulimova GE, Udina IG, Kunizheva SS, Kompaniyt-sev AA: Creating STS markers for human chromosome 3. Mol Biol (Mosk) 33:698–703 (1999).

Wang JY, Zabarovsky ER, Berglund P, et al: Somatic cell hybrid panel and *Not*I linking clones for physical mapping of the human chromosome 3. Genomics 20:105–113 (1994).

Wei MH, Latif F, Bader S, et al: Construction of a 600-kilobase cosmid clone contig and generation of a transcriptional map surrounding the lung cancer tumor suppressor gene (TSG) locus on human chromosome 3p21.3: progress toward the isolation of a lung cancer TSG. Cancer Res 56:1487–1492 (1996).

Venter JC, Adams MD, Myers EW, et al: The Sequence of the human genome. Science 291:1304–1351 (2001).

Zabarovsky ER, Kashuba VI, Pokrovskaya ES, et al: Alu-PCR approach to isolate *Not*I linking clones from the 3p14→p21 region frequently deleted in renal cell carcinoma. Genomics 16:713–719 (1993).

Zabarovsky ER, Allikmets R, Kholodnyuk I, et al: Construction of representative *Not*I linking libraries specific for the total human genome and for human chromosome 3. Genomics 20:312–316 (1994a).

Zabarovsky ER, Kashuba VI, Pettersson B, et al: Shot-gun sequencing strategy for long-range genome mapping: a pilot study. Genomics 21:495–500 (1994b).

Zabarovsky ER, Kashuba VI, Gizatullin RZ, et al: *Not*I jumping and linking clones as a tool for genome mapping and analysis of chromosome rearrangements in different tumors. Cancer Detect Prev 20:1–10 (1996).