

3. Ergebnisse

3.1. Isolation und Klonierung von *cd30*-Gen-Fragmenten

Analyse der Genbanken

Um vollständige Sequenzdaten des *cd30*-Gens zu erhalten, wurden zwei verschiedene, eine humane plazentare genomische Cosmid- und eine Lambda-Phagenbank mit Hilfe von [α^{32} P]-dCTP-markierten synthetischen Oligonukleotiden und cDNA-Fragmenten hybridisiert. Hierdurch wurden Klone gefunden werden, die fast das gesamte *cd30*-Gen abdecken. Für die Hybridisierung der Phagenbank wurde ein Oligonukleotid aus dem 5'-Bereich der CD30-cDNA benutzt. Um repräsentative Ergebnisse zu erhalten, wurden mehr als 10^6 Phagen- bzw. Cosmidklone aus jeder genomischen Bank untersucht.

Die erste Analyse der Genbanken ergab folgende Klone:

Klon	Herkunft	Größe des Inserts
M1	Lambda Fix II	~ 15 kB
H160	Cosmid	~ 40 kB
H180	Cosmid	~ 40 kB
G13	Cosmid	~ 40 kB
G40	Cosmid	~ 40 kB
G81	Cosmid	~ 40 kB

Die ungefähre Größe der DNA-Fragmente in den Klonen wurde durch Restriktionsenzymverdau mit selten schneidenden Enzymen, die in der Klonierungsschnittstelle des Vektors schneiden (z.B. NotI), und anschließende Gelelektrophorese bestimmt. Auf Grund der zum Hybridisieren benutzten Oligonukleotide bzw. des CD30-cDNA-Fragments (Acc.No.: M83554, Basenpaare 1 - 1905) aus dem 5'-Bereich ließen sich die beiden Klone M1 und H160 ohne weitere Analyse diesem Bereich des *cd30*-Gens zuordnen.

Analyse der Klone

Um vor Beginn der Sequenzierungsarbeiten festzustellen, welche Klone einen bestimmten Bereich des *cd30*-Gens abdecken, wurden denaturierte Proben der verschiedenen Klone mit unterschiedlichen [³²P]-markierten Oligonukleotiden aus verschiedenen Bereichen des *cd30*-Gens hybridisiert.

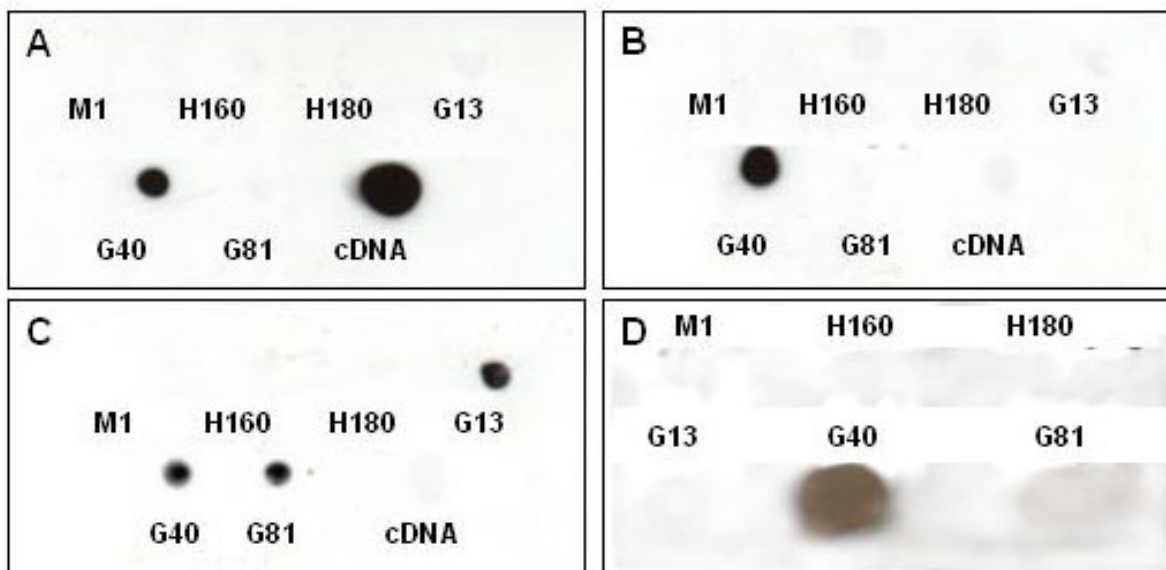


Abbildung 1: Autoradiographie der denaturierten DNA verschiedener Klone: Abb. A wurde mit dem Oligonukleotid 5'-GGGCCAGTGCTCTCTGGGTGATCCTGGTGTGG, Pos. +8928 – +8961 (Zugriffsnummer AJ289159, relativ zum Haupttranskriptionsstartpunkt) im 4. Exon, Abb. B mit dem Oligonukleotid 5'-GGGCGAGTGCATCATGCCTGGCTAATTTTTTGT, Pos. +16262 – +16295, Abb. C mit dem Oligonukleotid 5'-CTAAGTACAAGTGGGCTGGTTCTTAACTCCTCAA, Pos. +18548 – +18581 (3'-Bereich des Gens) und Abb. D mit dem Oligonukleotid 5'-GCTCAGATACTTATTGATGGTATAACCCTGGGCC, Pos. +14323 – +14356, hybridisiert.

Hieraus läßt sich die ungefähre Lage der Klone ableiten. Der Klon G40 besitzt die größte Ausdehnung und überlappt teilweise mit den Klonen G13 und G81. Da Oligonukleotide aus dem 3' Bereich des *cd30*-Gens verwendet wurden, läßt sich die Aussage bestätigen, daß die Klone M1 und H160 den 3'-Bereich des *cd30*-Gens nicht mit abdecken. Abbildung 2 zeigt dies als graphische Darstellung.

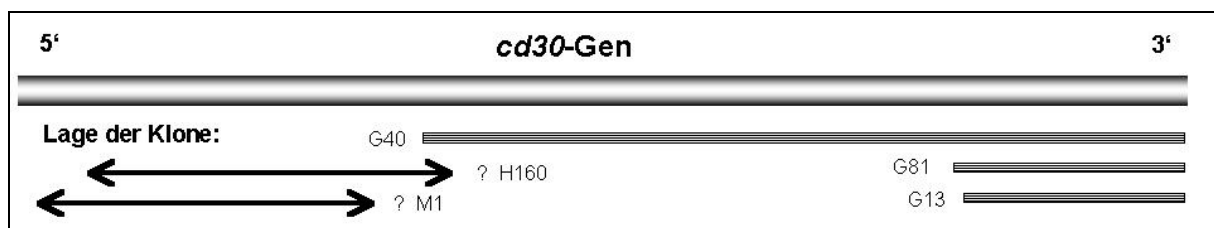


Abbildung 2: Die Klone G40, G13 und G81 decken den 3'-Bereich des *cd30*-Gens ab, M1 und H160 den 5'-Bereich.

Die gefundenen Klone wurden durch Restriktionsverdau und Gelelektrophorese weiter analysiert. Anhand identischer Fragmente konnten gleiche Klone identifiziert werden. Bei der Wahl der Restriktionsenzyme wurden solche mit mittlerer bis seltener Schnitzzahl benutzt (SmaI, EcoRI, SstI, KpnI, PstI), um eine geeignete Zahl an Fragmenten für die spätere Klonierung in den Vektor pBlueScript zu erhalten. Außerdem wurde ein Enzym benutzt, das im Bereich der Klonierungsschnittstelle des Vektors schneidet (NotI, SstI), um die Vektor-DNA von der humanen genomischen DNA abzutrennen.

Durch Anfertigen eines Southern Blots und anschließende Hybridisierung mit verschiedenen Oligonukleotiden konnten jene Fragmente identifiziert werden, die Fragmente des *cd30*-Gens enthielten. Abbildung 3 zeigt zwei Southern Blots, Teil A die Klone G40 und G13, die mit verschiedenen Restriktionsenzymen verdaut und mit einem CD30-cDNA-Fragment aus dem 3'-Bereich (Basenpaare 1151 – 3396, Acc.No. M83554) hybridisiert wurden. Das Muster aus mehreren Banden kommt dadurch zustande, daß die verwendete cDNA-Sonde einen Bereich abdeckt, der nach dem Restriktionsenzymverdau auf mehrere Fragmente verteilt ist. Die Klone G13 und G40 überlappen teilweise, wie an den gleichgroßen Fragmenten erkennbar ist. Teil B zeigt den Southern Blot mit dem Klon M1, der mit einem Oligonukleotid im 5'-Bereich hybridisiert wurde. Die Autoradiographie des hybridisierten Southern Blots des Klons H160 ergab größenidentische Fragmente, bei der Sequenzierung konnte bestätigt werden, daß die Klone M1 und H160 im 5'-Bereich des *cd30*-Gens überlappen.

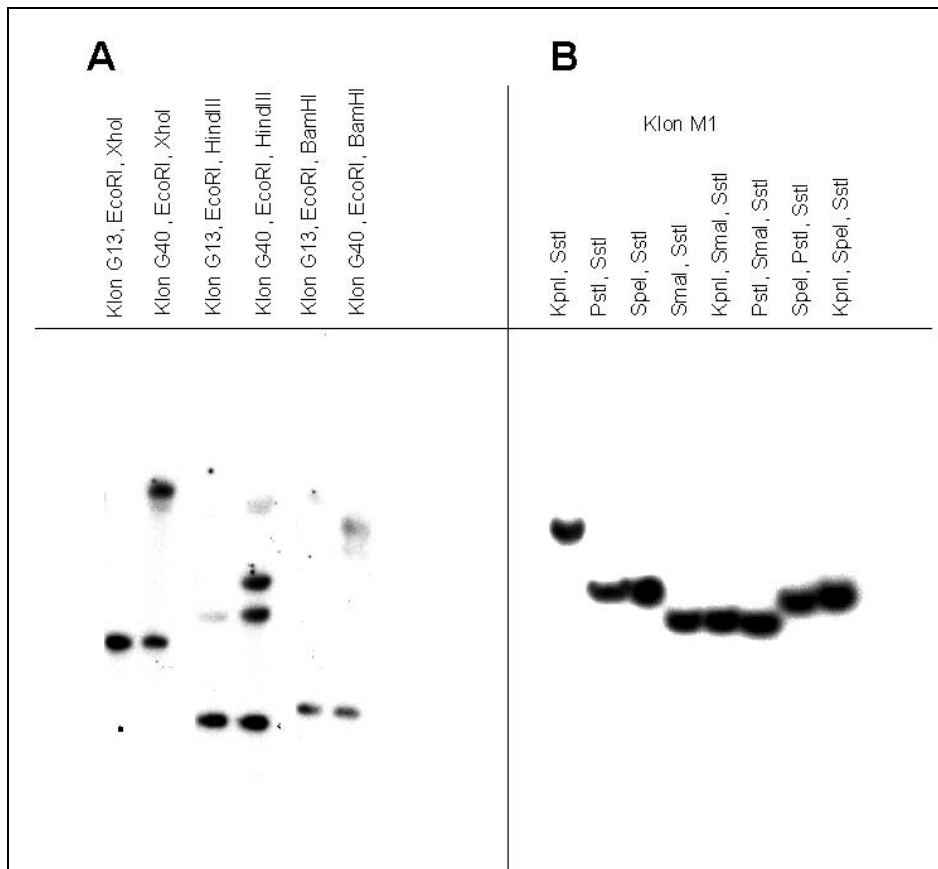


Abbildung 3: [A] zeigt den Southern Blot mit den Klonen G13 und G40, hybridisiert mit einem CD30-cDNA-Fragment aus dem 3'-Bereich, in [B] ist die Autoradiographie des Klons M1 dargestellt, der mit dem Oligonukleotid 5'-CCGATCCGCGTCCGGCAAAGCAACGCTGTGGA, Position +904 - +935, hybridisiert wurde. Die Autoradiographie des Klons H160 zeigte größenidentische Fragmente (nicht gezeigt).

Die restriktionsverdauten und gelelektrophoretisch aufgetrennten DNA-Fragmente eines Klons wurden nach ihrer Isolierung aus dem Gel in den pBlueScript-Vektor kloniert, so daß der aus der Genbank erhaltene Klon mit Gen-Fragmenten von 15 bis 40 kB in kleinere Fragmente zerlegt wurde. Die Verwendung verschiedener Restriktionsenzyme stellt dabei sicher, daß bei der Sequenzierung der verschiedenen *cd30*-Gen-Fragmente eine fortlaufende Sequenz durch überlappende Fragmente gewonnen werden kann.

3.2. Sequenzierung der *cd30*-Gen-DNA-Fragmente

Die Sequenzierung der einzelnen *cd30*-Gen-Fragmente wurde mit den im pBlueScript Vektor vorhandenen T7- und Sp6-Primern begonnen, nach Auswertung der Se-

quenzdaten wurden im 3'-Bereich der neuen Sequenz neue Oligonukleotide für die nächste Sequenzierungsreaktion ausgewählt. Die so erhaltenen Sequenzstücke wurden computergestützt zusammengesetzt. Um valide Sequenzdaten zu erhalten, wurde jedes Sequenzfragment vom 5'- und 3'-Ende aus sequenziert. Jede Sequenz wurde weiterhin mindestens dreimal durchsequenziert. Die einzelnen DNA-Fragmente mußten mindestens um 100 bp überlappen, damit ein sinnvolles Zusammensetzen der Sequenzen möglich war. Entsprechend wurde dies auch bei der Auswahl neuer Oligonukleotide zur Sequenzierung berücksichtigt.

Im Verlauf der Sequenzierung der DNA-Fragmente zeigte sich, daß Teile des 1. Introns mit dem 5'-Ende des 2. Exons noch nicht erfasst worden waren, dieser Sequenzbereich wurde mit Hilfe der Langstrecken-PCR amplifiziert. Im Bereich des 5. Intron zeigten sich bei der manuellen Sequenzanalyse Exon/Intron-Grenzen, die bei der computergestützten Analyse nicht gefunden worden waren, so daß auch dieser Teil mit Hilfe der Langstrecken-PCR (LR-PCR) nochmals amplifiziert und durch Sequenzierung verifiziert wurde.

Bei der Langstrecken-PCR konnten 4 PCR-Produkte gewonnen werden, die in den pCRII-TOPO Vektor kloniert wurden. Im Bereich des 1. Intron mit dem 5'-Ende des 2. Exons ergab die PCR die 3 Produkte, die wie folgt benannt wurden: 2B-ScaI (~7 kbp), 3B-DraI (~1.5 kbp) und 5B-SspI (~1.6 kbp). Im Bereich des 5. Introns ein Produkt: 15B-SspI (~900 bp) (s. Tabelle). Diese PCR-Produkte wurden ebenfalls sequenziert und in die bereits vorhandene Sequenz integriert.

Klon	Herkunft	Größe des PCR-Produkts
2B-Sca I (Intron 1)	LR-PCR	~ 7 kbp
3B-Dra I (Intron 1)	LR-PCR	~ 1.5 kbp
5B-Ssp I (Intron 1)	LR-PCR	~ 1.6 kbp
15B-Ssp I (Intron 5)	LR-PCR	~ 900 bp

Nachdem alle DNA-Fragmente sequenziert und zu einer Konsensus-Sequenz zusammengesetzt worden waren, wurde diese mit der CD30-cDNA-Sequenz durch das Programm MegAlign (DNASstar) bzw. Bestfit (HUSAR) abgeglichen, um sicherzustellen, daß das ganze *cd30*-Gen erfasst wurde. Hier zeigte sich, daß die Konsen-

sten drei cysteinreichen Domänen des extrazellulären Rezeptorteils, das 3. Exon die letzten drei cysteinreichen Domänen.

Die Transmembrandomäne des CD30-Rezeptors wird von den letzten Basenpaaren des 3. Exons sowie hauptsächlich vom 4. Exon codiert. Die Transmembrandomäne besteht aus einer 27 Aminosäuren langen hydrophoben Region, die jeweils von geladenen, hydrophilen Aminosäuren begrenzt wird.

Der intrazelluläre Teil des CD30 Rezeptors befindet sich in den Exonen 5 bis 8, wobei die für die Signaltransduktion notwendigen Domänen im 8. Exon liegen (PEQET, EEEGKE, EPPLGSC Aminosäuresequenzen). Ein großer Teil des 8. Exons beinhaltet die nicht-translatierte Region der CD30 mRNA.

3.4. Promoterregion des *cd30*-Gens

Eine ca. 1000 bp lange Sequenz im 5'-Bereich des *cd30*-Gens (5'-Bereich bis Anfang des 1. Exons) wurde auf das Vorkommen von möglichen Bindungsstellen für Transkriptionsfaktoren analysiert. Hierbei zeigte sich, daß im CD30-Promoter keine TATA oder CAAT Box vorhanden ist. Eine dem *downstream promoter element* (DPE)^{35,36} ähnliche Sequenz (^A/G^A/T^TCGTG) konnte bei Position (ab angegebenem Transkriptionsstart, Acc.-Nr.: AJ289159) +20 bis +26 bp (GCACGTG) identifiziert werden. Eine überwiegend der Initiator (Inr) - Konsensussequenz³⁷ (PyPyAN^T/A^APyPy) entsprechende Sequenz (TCAAACA) findet sich bei Position -64 bis -58 (Zugriffsnummer AJ 289159) relativ zum Haupttranskriptionsstartpunkt.

Die wahrscheinliche Promoterregion enthaltende Sequenz (s.o.) wurde mit Hilfe des Programms FACTOR des HUSAR-Programmpakets nach möglichen Transkriptionsfaktorbindungsstellen durchsucht. In diesem Bereich fand sich eine sehr große Anzahl von möglichen Transkriptionsfaktorbindungsstellen, auffällig waren 4 GC-reiche Konsensus-Bindungsstellen, die Sp1- oder Sp3-Bindungsstellen entsprechen könnten.

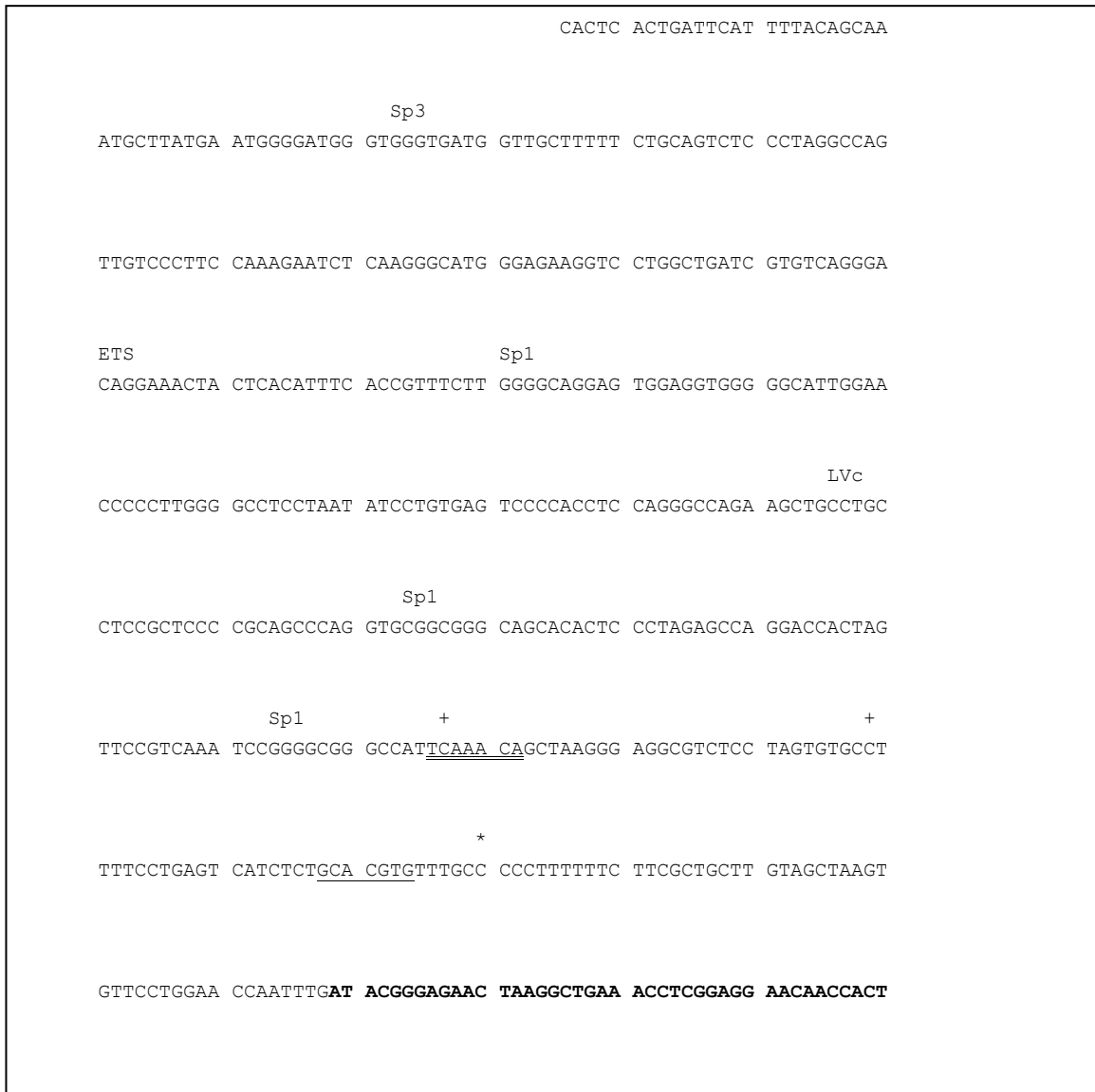


Abbildung 4 Wahrscheinliche Promoterregion des CD30. Der Beginn der cDNA ist **fett** gedruckt, das *DPE* ist unterstrichen, die Inr-Sequenz ist doppelt unterstrichen. Weiterhin sind Bindungsstellen für Transkriptionsfaktoren gezeigt, ermittelt mit Hilfe des Programms FACTOR des HUSAR Softwarepakets. Diese Bindungsstellen sind funktionell aktiv, wie in weiteren Untersuchungen innerhalb unserer Arbeitsgruppe gezeigt werden konnte. Ebenso sind die verschiedenen Transkriptionsstartpunkte angegeben, * ist der Haupttranskriptionsstartpunkt in HL-Zelllinien, die beiden anderen Transkriptionsstartpunkte sind durch ein + markiert.

Diese Untersuchungen zeigen, daß das *cd30*-Gen durch einen Promoter ohne TATA oder CAAT-Box reguliert wird. Die Sequenzanalyse ergab funktionelle Elemente, wie sie für einen TATA-losen Promoter beschrieben worden sind, so die tandemartig aufgereihten Bindungsstellen für Sp1-Transkriptionsfaktoren, die einen Transkriptionsstartpunkt in diesem Bereich nahelegen.

3.5. Mikrosatelliten-Sequenz

In der 5'-Region des *cd30*-Gens fand sich ca. 700 bp vor dem 1. Exon (-697 bis -427, Acc.No. 289159) eine ca. 200 bp lange Sequenz von ATCC Wiederholungen. Diese Mikrosatellitenstruktur wurde in unterschiedlichen Zelllinien und Geweben auf einen möglichen genetischen Polymorphismus hin untersucht. Hierzu wurde diese Region mit Hilfe der Langstrecken-PCR aus genomischer DNA der CD30⁺ HL-(Hodgkin-Lymphom)-Zelllinien Ho (EMBL Acc.Nr.: AJ308521), L428 (AJ311490), L591 (AJ311493), L1236 (AJ311488) und L540 (AJ311494) sowie der ALCL-Zelllinien DEL (AJ311491) und SU-DHL1 (AJ311492) amplifiziert. Zum Vergleich wurde dieselbe Region aus nicht-neoplastischem humanem Gewebe (Pankreas (AJ311487), 3 Plazenten (AJ308522, AJ408011, AJ409012) und Tonsille (AJ311489)) amplifiziert (s. Abbildung 4).

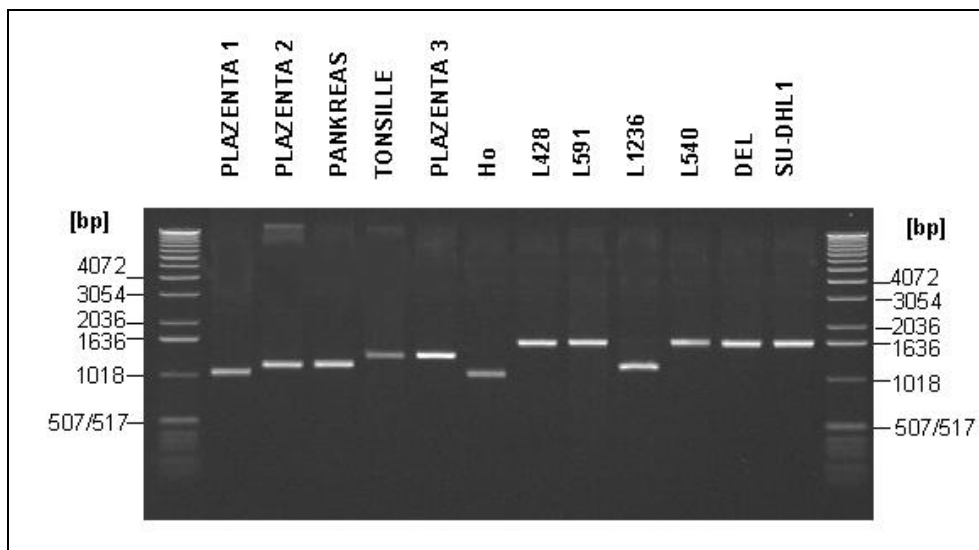


Abbildung 4: Agarosegelbild der amplifizierten PCR Produkte der Mikrosatellitenregion verschiedener Zelllinien. Deutlich erkennbar ist die unterschiedliche Größe der Banden als Ausdruck eines Längenpolymorphismus der ATCC-Mikrosatelliten.

In allen durchgeführten PCR-Reaktionen wurde ein einzelnes PCR-Produkt generiert. Auch unter geänderten PCR-Bedingungen und mit verschiedenen Primern ließ sich kein zweites PCR-Produkt als Ausdruck eines zweiten Allels amplifizieren. Die PCR-Produkte unterschieden sich in der Länge um maximal 500 bp mit einer

Produktlänge zwischen 1050 bp bis 1612 bp (s. Abbildung 4). Die erhaltenen PCR-Produkte wurden in den pCRII-TOPO Vektor kloniert und sequenziert.

In der Sequenzanalyse zeigte sich eine unterschiedliche Anzahl von ATCC Wiederholungen als Ursache für die unterschiedlichen Produktlängen. Es fanden sich weiterhin Varianten des ATCC Motivs (ACCC, ATGC, GTCC), die in allen PCR-Produkten sporadisch und unabhängig von dem untersuchten PCR-Produkt vorkamen. Die Untersuchung der Sequenzen auf mögliche Transkriptionsfaktorbindungsstellen ergab durch den Polymorphismus neu entstandene mögliche Bindungsstellen in einigen PCR-Produkten. Hier fanden sich neben Konsensussequenzen für weitere Transkriptionsfaktoren vor allem zusätzliche Glucocorticoid-Repressor-Elemente und Bindungsstellen für die NF-1 Transkriptionsfaktor-Familie.

Einen schematischen Überblick der Promoterregion des *cd30*-Gens mit den verschiedenen Elementen gibt die folgende Abbildung:

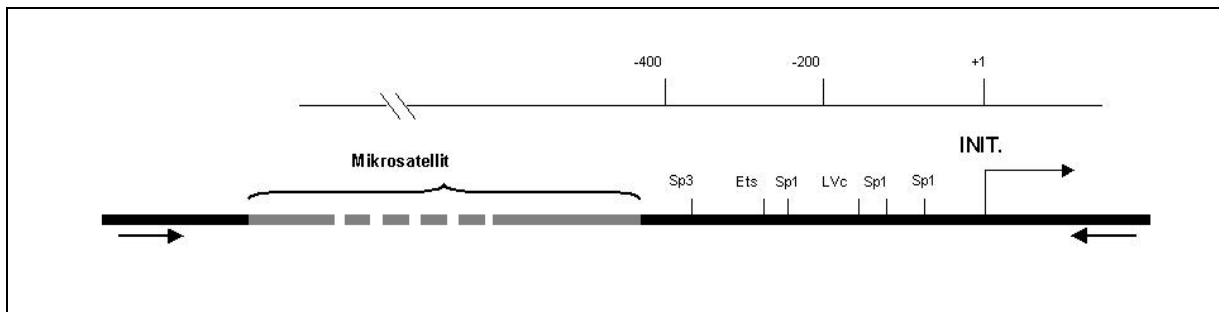


Abbildung 5: Schematischer Überblick über die Promoterregion des *cd30*-Gens. Dargestellt ist die Mikrosatellitenregion und die nachfolgende Region mit den Bindungsstellen für verschiedene Transkriptionsfaktoren. Der Transkriptionsstartpunkt ist ebenfalls angegeben (Init). In der obersten Zeile ist der Größenmaßstab. Die für die PCR benutzten Primer sind als Pfeile in der untersten Zeile eingetragen.

3.6. Karte des *cd30*-Gens

Nach den vorausgegangenen Analysen ließ sich folgende graphische Darstellung des *cd30*-Gens mit seiner Exon/Intron-Aufteilung und den die entsprechenden Regionen abdeckenden Klonen erstellen:

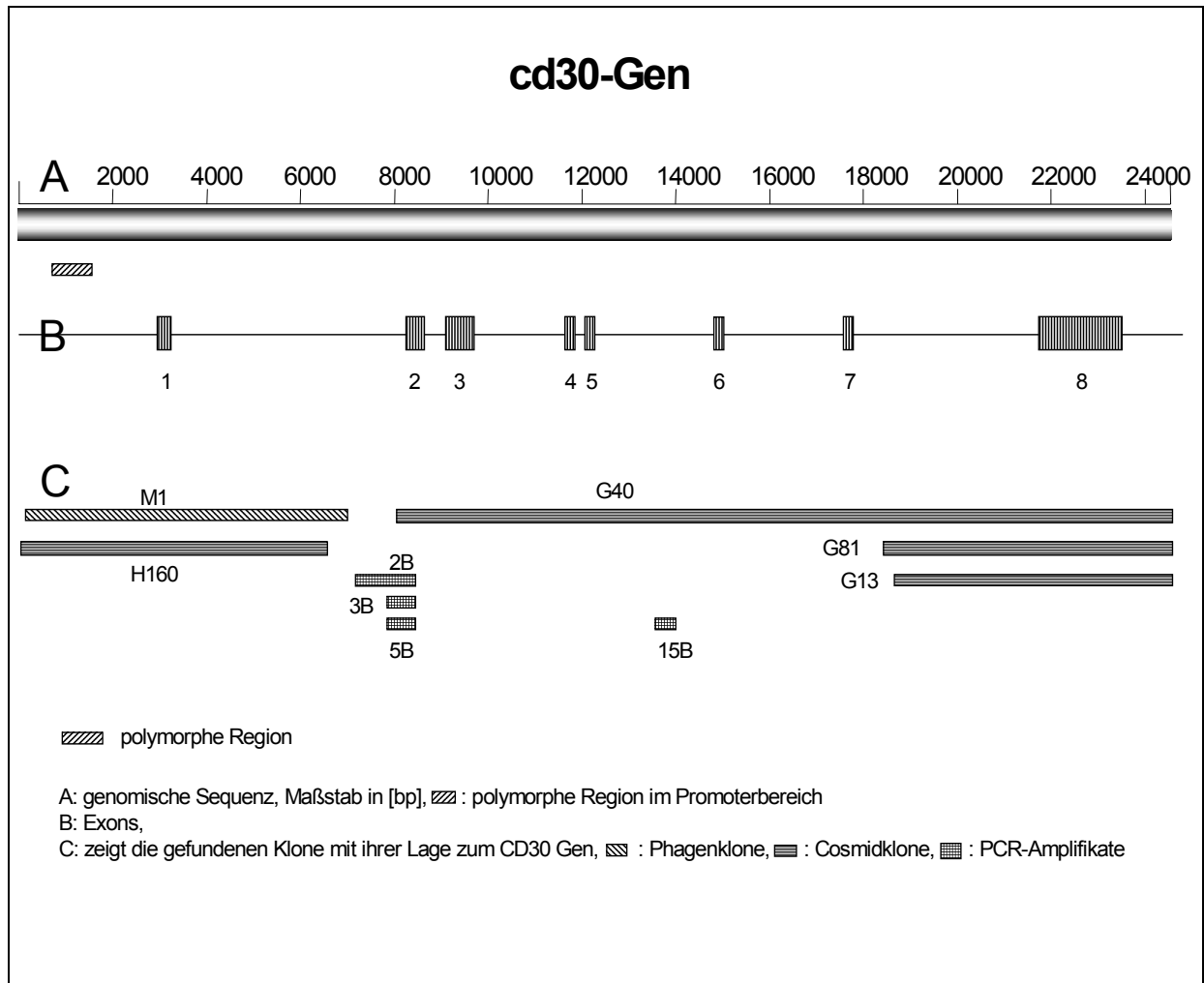


Abbildung 6: In Teil A ist der Größenmaßstab sowie die polymorphe Region eingezeichnet, Teil B zeigt die Anordnung der Exonen und Intronen Teil C die Lage der Klonen und PCR Amplifikate anhand derer das *cd30*-Gen sequenziert wurde.

3.7. Vergleich des *cd30*-Gens mit der TNFR-Superfamilie

Die Korrelation der Genstruktur mit der Peptidsequenz weist beim *cd30*-Gen im Vergleich mit den anderen Mitgliedern der TNF-Rezeptor-Superfamilie einige Unterschiede auf. Bei den meisten Mitgliedern der TNFR Superfamilie, wie zum Beispiel TNFR1³⁸, TNFR2³⁹, CD27⁴⁰ und CD95⁴¹, sind innerhalb der cysteinreichen Region konservierte Splice-Stellen vorhanden, die die einzelnen konservierten, jeweils 6 Cysteinreste und ca. 40 Aminosäuren umfassenden Cysteinreichen-Motive unterbrechen, beim *cd30*-Gen sind jedoch keine Splice-Stellen innerhalb der einzelnen cysteinreichen Motive vorhanden. Es findet sich beim *cd30*-Gen nur eine Splice-Stelle zwischen den beiden aus jeweils 3 cysteinreichen Motiven aufgebauten Untereinheiten. Wie die Sequenzanalyse mit dem Programm MegAlign zeigt, sind diese Untereinheiten, entsprechend den Exonen 2 und 3, auf Nukleinsäureebene zu großen Teilen homolog (ca 72%).