



Roger Colobran i Oriol TESI DOCTORAL 2008

**Estudi Immunogenètic de les Quimiocines  
CCL4 i CCL4L: Exemple i Model de la  
Complexitat de la Superfamília de  
Quimiocines**

Roger Colobran i Oriol

2008

**Universitat Autònoma de Barcelona**  
**Facultat de Ciències**  
**Departament de Biologia Cel·lular, Fisiologia i Immunologia**

**Estudi Immunogenètic de les Quimiocines CCL4 i CCL4L:  
Exemple i Model de la Complexitat de la Superfamília de  
les Quimiocines**

Autor: Roger Colobran i Oriol

Direcció: Manel Juan i Otero

Memòria de la Tesi presentada per obtenir el grau de Doctor en Ciències Biològiques per a  
la Universitat Autònoma de Barcelona

Gener, 2008



IIGTiP BST LIRAD

*Crtra. del Canyet s/n,*  
*08916 Badalona*  
tel. 34.93.497.86.66  
fax. 34.93.497.86.68  
e-Mail [mjuan@clinic.ub.es](mailto:mjuan@clinic.ub.es)



Universitat  
Autònoma de  
Barcelona

Departament de Biologia Cel·lular,  
Fisiologia i Immunologia.  
Hospital Universitari  
Germans Trias i Pujol.  
Universitat Autònoma de  
Barcelona.

Laboratori d'Immunobiologia per a  
la Recerca i Aplicacions  
Diagnòstiques (LIRAD)  
**Banc de Sang i Teixits (BST)**

Manel Juan i Otero, Professor Associat d'Immunologia de la Universitat Autònoma de Barcelona i Metge Especialista en Immunologia al LIRAD,

Certifica:

Que la tesi titulada "***Estudi Immunogenètic de les Quimiocines CCL4 i CCL4L: Exemple i Model de la Complexitat de la Superfamília de les Quimiocines***" ha estat realitzada per Roger Colobran i Oriol sota la seva direcció i considera que és apta per a ésser presentada i optar al grau de Doctor en Ciències Biològiques per a la Universitat Autònoma de Barcelona.

I per tal que en quedi constància signa la present certificació a Badalona 20 de Desembre 2007.

Dr. Manel Juan i Otero

## ABREVIATURES

CAMs	<i>Cell-Adhesion Molecules</i>
GlyCAM-1	<i>Glycosylation-dependent Cell Adhesion Molecule-1</i>
PSGL-1	<i>P-Selectin Glycoprotein Ligand-1</i>
ICAMs	<i>InterCellular Adhesion Molecules</i>
VCAM-1	<i>Vascular Cell Adhesion Molecule-1</i>
MAdCAM-1	<i>Mucosal Addressin Cell Adhesion Molecule-1</i>
VAPs	<i>Vascular Adhesion Proteins</i>
HEVs	<i>High Endothelial Venules</i>
MIP	<i>Macrophage Inflammatory Protein</i>
PAF	<i>Platelet-Activating Factor</i>
VAs	<i>Vascular Addressins</i>
HCC	<i>Hemofiltrate CC chemokine</i>
MPIF	<i>Myeloid Progenitor Inhibitory Factor</i>
RANTES	<i>Regulated on Activation, Normal T cell Expressed and Secreted</i>
GAGs	<i>Glycosaminoglycans</i>
GCP	<i>Granulocyte Chemotactic Protein</i>
MIG	<i>Monokine Induced by IFN<math>\gamma</math></i>
IP-10	<i>IFN<math>\gamma</math>--Inducible Protein 10</i>
MCP	<i>Monocyte Chemotactic Protein</i>
I-TAC	<i>IFN<math>\gamma</math>-Inducible T cell <math>\alpha</math> Chemoattractant</i>
SDF	<i>Stromal cell-Derived Factor</i>
MDC	<i>Macrophage-Derived Chemokine</i>
MSGA/GRO	<i>Melanoma Growth Stimulatory Activity / GROwth regulated gene</i>
NAP	<i>Neutrophil Activating Peptide</i>
ENA	<i>Epithelial cell-derived Neutrophil Attractant</i>
PF4	<i>Platelet Factor 4</i>
BLC/BCA	<i>B Lymphocyte Chemoattractant / B Cell-Attracting chemokine</i>
PARC	<i>Pulmonary and Activation-Regulated Chemokine</i>
TECK	<i>Thymus-Expressed ChemoKine</i>
TARC	<i>Thymus and Activation Regulated Chemokine</i>
SLC	<i>Secondary Lymphoid tissue Chemokine</i>
ELC	<i>EBI1-Ligand Chemokine</i>
LARC	<i>Liver- and Activation-Regulated Chemokine</i>
CNV	<i>Copy Number Variation</i>

***ÍNDEX***

<b>1.</b>	<b>INTRODUCCIÓ</b> .....	<b>1</b>
<b>1.1.</b>	<b>Sistema immunitari: recirculació i inflamació</b> .....	<b>3</b>
1.1.1.	<i>El sistema immunitari</i> .....	3
1.1.2.	<i>La recirculació leucocitària</i> .....	5
1.1.3.	<i>Les molècules d'adhesió cel·lular</i> .....	5
1.1.4.	<i>L'extravassació dels neutròfils</i> .....	7
1.1.5.	<i>L'extravassació dels limfòcits</i> .....	9
<b>1.2.</b>	<b>Les quimiocines</b> .....	<b>10</b>
1.2.1.	<i>Classificació</i> .....	11
1.2.2.	<i>Estructura i relacions estructura-funció de les quimiocines</i> .....	12
1.2.3.	<i>Classificació funcional de les quimiocines CXC</i> .....	14
1.2.3.1.	<i>Funció inflamatòria de les quimiocines CXC (ELR<sup>+</sup>)</i> .....	14
1.2.3.2.	<i>Altres funcions vasculars de les quimiocines CXC</i> .....	15
1.2.3.3.	<i>Funció homeostàtica de les quimiocines CXC</i> .....	16
1.2.4.	<i>Funcions de les quimiocines CC</i> .....	17
1.2.4.1.	<i>Grup de quimiocines CC proinflamatòries</i> .....	17
1.2.4.2.	<i>Grup de quimiocines CC homeostàtiques i del desenvolupament</i> .....	19
<b>1.3.</b>	<b>Els receptors de quimiocines</b> .....	<b>21</b>
<b>1.4.</b>	<b>La quimiocina CCL4 (MIP-1<math>\beta</math>)</b> .....	<b>23</b>
1.4.1.	<i>Noms alternatius</i> .....	23
1.4.2.	<i>Descripció estructural de la proteïna</i> .....	23
1.4.3.	<i>Principals cèl·lules i teixits que expressen CCL4</i> .....	24
1.4.4.	<i>Receptors</i> .....	24
1.4.5.	<i>Funcions principals</i> .....	24
1.4.6.	<i>Intervenció en patologia</i> .....	25
1.4.7.	<i>Localització cromosòmica i altres aspectes importants</i> .....	25

<b>1.5. Variació genòmica</b>	<b>28</b>
1.5.1. <i>Glossari general</i>	28
1.5.1.1. SNPs	28
1.5.1.2. Repeticions nucleotídiques	28
1.5.1.3. Altres elements de variabilitat genòmica: Insercions/Delecions, Inversions i Translocacions	29
1.5.2. <i>Variants estructurals</i>	30
<b>2. HIPÒTESI</b>	<b>33</b>
<b>3. OBJECTIUS</b>	<b>37</b>
<b>4. RESULTATS</b>	<b>41</b>
<b>4.1. Capítol I</b>	<b>43</b>
4.1.1. <i>Consideracions prèvies sobre l'article</i>	45
4.1.2. <i>Resum de dades del primer capítol</i>	45
<b>4.2. Capítol II</b>	<b>57</b>
4.2.1. <i>Consideracions prèvies sobre l'article</i>	59
4.2.2. <i>Resum de dades del segon capítol</i>	59
<b>4.3. Capítol III</b>	<b>67</b>
4.3.1. <i>Consideracions prèvies sobre l'article</i>	69
4.3.2. <i>Resum de dades del tercer capítol</i>	69
<b>4.4. Capítol IV</b>	<b>109</b>
4.4.1. <i>Consideracions prèvies sobre l'article</i>	111
4.4.2. <i>Resum de dades del quart capítol</i>	111

---

<b>4.5. Capítol V .....</b>	<b>123</b>
<i>4.5.1. Consideracions prèvies sobre l'article .....</i>	125
<i>4.5.2. Resum de dades del cinquè capítol .....</i>	125
<b>5. DISCUSSIÓ .....</b>	<b>139</b>
<b>5.1 Discussió del capítol I .....</b>	<b>141</b>
<b>5.2 Discussió del capítol II .....</b>	<b>145</b>
<b>5.3 Discussió del capítol III .....</b>	<b>146</b>
<b>5.4 Discussió dels capítols IV i V .....</b>	<b>149</b>
<b>6. CONCLUSIONS .....</b>	<b>151</b>
<b>7. REFERÈNCIES .....</b>	<b>155</b>





# ***1. INTRODUCCIÓ***



# 1. INTRODUCCIÓ

## 1.1. SISTEMA IMMUNITARI: RECIRCULACIÓ I INFLAMACIÓ

### 1.1.1. El sistema immunitari

El sistema immunitari (S.I.) és un sistema de reconeixement perfectament adaptat que ha evolucionat en els vertebrats per protegir-los de les infeccions (i també del càncer). Per desenvolupar aquesta funció, el S.I. genera una enorme varietat de cèl·lules i molècules capaces de reconèixer específicament i eliminar una varietat aparentment il·limitada d'invasors exteriors.

Funcionalment, la resposta immunitària pot dividir-se en dues etapes relacionades: *reconeixement* i *resposta efectora*. El **reconeixement** es caracteritza per la seva especificitat. El sistema immunitari és capaç de reconèixer diferències químiques subtils que distingeixen un determinat patògen d'un altre. A més a més, és capaç de discriminar entre allò "propi" i allò "no propi", tot i que aquesta simplificació conceptual ha sigut replantejada, en part, per certs immunòlegs que prefereixen afirmar que el sistema immunitari és capaç de distingir entre allò "perillós" i allò "no perillós" (sovint hi ha absència de resposta en front a elements no propis que no representen perill).

Quan un determinat microorganisme és reconegut, el sistema immunitari mobilitza una gran varietat de cèl·lules i molècules que participaran en la formació d'una resposta apropiada (**resposta efectora**) destinada a eliminar o neutralitzar l'organisme.

La resposta immunitària (i per extensió els elements que hi participen) es poden dividir en el que anomenem immunitat innata i immunitat adaptativa. La **immunitat innata** inclou els mecanismes no específics de la lluita contra els patògens; la formen essencialment les barreres físiques i químiques (com epitelis i substàncies antimicrobianes), certes proteïnes sanguínies (entre les que s'inclouen membres del sistema del complement, citocines i altres mediadors de la inflamació), cèl·lules fagocítiques (neutròfils, macròfags) i altres leucòcits (com són les cèl·lules NK), constituint la primera línia de defensa contra els microorganismes. La **immunitat adaptativa** mostra un elevat grau d'**especificitat** (per això sovint també se l'anomena immunitat específica), la qual cosa permet respondre de forma singular i eficient contra diferents tipus de microorganismes. Una de les característiques més rellevants de la immunitat adaptativa, a part de la seva especificitat, és que "recorda" totes les trobades amb un microorganisme o antígen estrany, de tal forma que contactes posteriors estimulen mecanismes de defensa cada cop més eficaços. Aquesta propietat s'anomena

**memòria immunològica** i constitueix el fonament de la vacunació protectora en contra les malalties infeccioses. Els components cel·lulars principals de la immunitat específica són els limfòcits, i una de les característiques distintives d'aquestes cèl·lules explica la tercera de les característiques de la immunitat adaptativa, la **clonalitat**, doncs cada cèl·lula té un receptor específic que possibilita que, davant d'un únic estímul antigènic, la resposta es desenvolupi a partir de l'expansió i diferenciació d'un reduït nombre de cèl·lules. La diversitat funcional i estructural dels propis limfòcits i els seus productes (d'entre els que destaquen els **anticossos**) és el que al final defineix les peculiaritats (resposta humoral i cel·lular) d'aquesta resposta adaptativa. Tot i aquesta diferenciació, els mecanismes de les respostes immunitàries innata i específica formen un sistema integrat de defensa en el que existeix una cooperació funcional estructurada de nombroses cèl·lules i molècules (de fet la resposta adaptativa es basa en la resposta natural i no existeix sense aquesta).

Amb la finalitat d'optimitzar aquestes interaccions cel·lulars, necessàries per les fases de reconeixement i activació de la resposta immunitària específica, la majoria de les cèl·lules implicades (**limfòcits, fagòcits mononuclears** i altres **cèl·lules accessòries**), es localitzen, concentren i structuren de manera ordenada en òrgans o teixits anatòmicament definits, que són també els llocs on es transporten i concentren els antígens estranys. Aquesta compartimentalització anatòmica no és fixa perquè, com veurem posteriorment, molts limfòcits recirculen i intercanvien constantment el seu lloc entre la circulació i els teixits. De fet, cal tenir en compte que el sistema immunitari és un sistema dispers per tot l'organisme, però que manté una unitat funcional gràcies als complexos mecanismes d'organització i relació entre tots els seus components. La mobilitat i la localització regulada dinàmicament de les cèl·lules del sistema immunitari és l'única manera de garantir una resposta immunitària eficient quan hi ha un reconeixement clonal, que significa una especificitat de pocs components. Sense aquesta mobilitat regulada no podria haver-hi resposta clonal.

Per la seva funció, els òrgans i teixits del sistema immunitari poden distingir-se en dos grups: a) els **teixits limfoides primaris** (també anomenats centrals o generadors) són els teixits en els que els limfòcits expressen per primera vegada els receptors antigènics i, en definitiva, adquireixen la maduresa fenotípica i funcional, i b) els **teixits limfoides secundaris** (o perifèrics) són els llocs on s'inicia i es desenvolupa la resposta dels limfòcits als antígens estranys. Els òrgans limfoides primaris dels mamífers són la **medul·la òssia** (d'on procedeixen tots els limfòcits i on maduren els limfòcits B), i el **timus** (on les cèl·lules T maduren i aconseguen l'estat de competència funcional). Els teixits limfoides perifèrics inclouen principalment els **ganglis limfàtics**, la **melsa**, el

**teixit limfàtic associat a mucoses (MALT) i el sistema immunitari de la pell.** A més a més, trobem agregats mal definits de limfòcits en els teixits connectius i en quasi tots els òrgans.

### 1.1.2. La recirculació leucocitària

Com hem acabat de comentar, el sistema immunitari no està localitzat en un determinat compartiment anatòmic. Per la pròpia funció de protecció, presenta un gran nombre d'òrgans i teixits morfològicament i funcional diferents que es troben repartits per la totalitat de l'organisme. A més a més, els elements efectors de les respostes (tant innata com específica) són "elements mòbils" que es desplacen per tot l'organisme a través del sistema sanguini i del sistema limfàtic.

Per la pròpia organització anatòmica dels òrgans i teixits del sistema immunitari, molts tipus de leucòcits es mouen d'una part del cos a una altra d'una manera perfectament regulada. Això és especialment cert en el cas dels **limfòcits**, ja que circulen contínuament a través de la sang i la limfa en direcció als diferents òrgans limfoides, des d'on surten per tornar a la circulació sanguínia i limfàtica. El procés de contínua **recirculació limfocitària** (fisiologia) no només permet augmentar al màxim la possibilitat de que els limfòcits interaccionin amb els antígens que es troben concentrats en els teixits limfoides, iniciant d'aquesta forma una resposta immunitària, sinó que també s'integra, juntament amb la participació d'altres elements cel·lulars efectors (principalment granulòcits i monòcits), en el desenvolupament de les **respostes inflamatòries** (patologia). La inflamació és una resposta complexa a un dany local que es caracteritza per la presència en el lloc de la lesió de vermellor, calor, inflor i, sovint, dolor. Aquests efectes clínics són el fruit d'un procés actiu que involucra diverses cèl·lules del sistema immunitari (natural i adaptatiu) i nombrosos mediadors. Cal però ressaltar que l'establiment i regulació de les respostes inflamatòries s'explica principalment per la migració controlada de poblacions leucocitàries al lloc de la lesió.

Existeixen diferents vies i mecanismes de recirculació limfocitària, però totes tenen en comú el desplaçament de cèl·lules a través de la sang i la limfa, i la posterior extravassació dels limfòcits dins els teixits limfoides secundaris o regions d'inflamació.

### 1.1.3. Les molècules d'adhesió cel·lular

El moviment dels leucòcits i de les molècules transportades per la sang cap a l'interior dels teixits es regula per l'endoteli vascular. Per tal que els leucòcits circulants puguin

entrar dins els teixits inflamats o dins els òrgans limfoides perifèrics, s'han d'adherir a les cèl·lules endotelials (que es troben tapissant les parets dels vasos sanguinis) i passar a través d'elles, en un procés anomenat **extravassació**. Les cèl·lules endotelials expressen **molècules d'adhesió cel·lular** (CAMs, de l'anglès *cell-adhesion molecules*) específiques de leucòcits. Algunes d'aquestes proteïnes de membrana s'expressen constitutivament; d'altres ho fan només en resposta a concentracions locals de citocines produïdes durant una resposta inflamatòria. Els limfòcits, monòcits i granulòcits presenten receptors que s'uneixen a les molècules d'adhesió presents a l'endoteli vascular, la qual cosa permet que aquestes cèl·lules s'extravassin dins els teixits. D'aquesta manera podem dir que en els processos de recirculació i inflamació la extravassació és cosa de dos components cel·lulars: els leucòcits i l'endoteli, que fan de "parella de ball" en aquestes situacions.

La majoria de les molècules d'adhesió cel·lular d'interès en immunologia pertanyen a quatre famílies de proteïnes:

a) **Família de les selectines**: Són glucoproteïnes de membrana que tenen a la part més distal un domini lectina que els permet interaccionar amb grups de carbohidrats específics. Inclou tres molècules designades com a L, E i P. Molts leucòcits circulants expressen la L-selectina, mentre que la E-selectina i la P-selectina són expressades bàsicament per les cèl·lules endotelials vasculares. Les selectines són responsables principals de la unió inicial dels leucòcits a l'endoteli vascular [1].

b) **Família de molècules tipus mucina**: Les mucines són un grup de proteïnes riques en serina i treonina altament glucosilades. La seva estructura els permet presentar-se com a lligands de les selectines. Per exemple, la L-selectina, present en els leucòcits, reconeix una estructura glucídica (sLe<sup>x</sup>) que es troba, entre d'altres, formant part de dues molècules altament glicosilades (CD34 i GlyCAM-1) expressades en certes cèl·lules endotelials dels ganglis limfàtics. Una altra molècula tipus mucina (PSGL-1), que es troba en neutròfils, interacciona amb la E- i P-selectines expressades en l'endoteli inflammat [2-4].

c) **Família de les integrines**: Són proteïnes heterodimèriques (formades per una cadena  $\alpha$  i una cadena  $\beta$ ) que expressen els leucòcits i faciliten tant l'adherència a l'endoteli vascular com altres interaccions cèl·lula - cèl·lula. A més a més, són elements centrals de la interacció cèl·lula - matriu extracel·lular. Les integrines s'agrupen en categories diferents segons quina subunitat  $\beta$  contenen. Diferents integrines són expressades per diferents poblacions de leucòcits, permetent a aquestes cèl·lules unir-se a diferents CAMs (pertanyents a la superfamília de les immunoglobulines) que s'expressen a l'endoteli vascular [5-7].

d) **Superfamília de les immunoglobulines:** Són molècules d'adhesió que contenen un nombre variable de dominis tipus immunoglobulina. En aquest grup (entre moltes d'altres molècules d'aquesta gran superfamília) s'inclouen ICAM-1, ICAM-2, ICAM-3 i VCAM-1, que són expressades per cèl·lules endotelials vasculares i s'uneixen a diverses integrines [8-10]. Una molècula d'adhesió anomenada MAdCAM-1 s'expressa a l'endoteli de mucoses i dirigeix l'entrada de limfòcits dins la mucosa. Presenta tant dominis tipus immunoglobulina (a través dels quals s'uneix a integrines) com dominis tipus mucina (unint-se a selectines a través d'ells) [11]. Així s'exemplifica que existeixen molècules classificables en un o altre grup.

Fora d'aquestes 4 famílies existeixen altres molècules d'adhesió com les cadherines (una família de glicoproteïnes de membrana que tenen en comú el fet de ser molècules d'adhesió dependents del calci) i d'altres no agrupables (com les VAPs, de l'anglès *Vascular Adhesion Protein*) que també són importants en determinades interaccions adhesives.

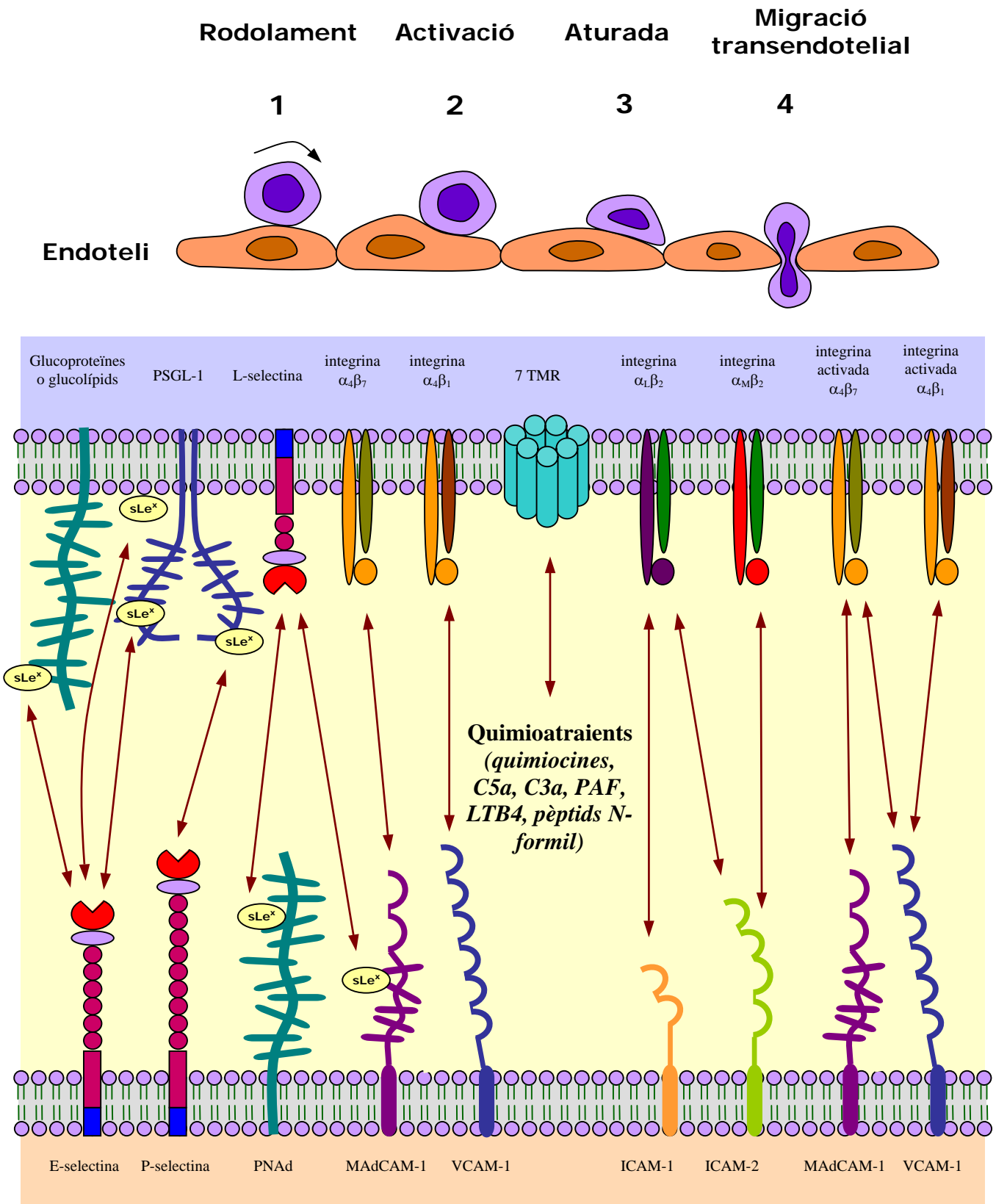
#### 1.1.4. L'extravassació dels neutròfils

Quan es desenvolupa una resposta inflamatòria, diverses citocines i altres mediadors inflamatoris actuen sobre els vasos sanguinis locals induint l'expressió de CAMs endotelials. És llavors quan es diu que l'endoteli està activat o inflammat. Dins el conjunt de cèl·lules que intervenen en el procés inflamatori, els neutròfils són, generalment, les primeres cèl·lules que s'uneixen a l'endoteli inflammat i s'extravassen dins els teixits. Per tant, els neutròfils han de reconèixer aquest endoteli inflammat, adherir-s'hi fortament (evitant ser arrossegats pel corrent sanguini) i penetrar, a través de la barrera endotelial, cap al teixit diana. Aquest procés d'extravassació dels neutròfils (i dels altres leucòcits) es pot dividir en quatre passos seqüencials: (1) **enganxament o lligament** (*tethering*) i **rodolament** (*rolling*), (2) **activació** (*activation*) **per estímuls quimioatracients**, (3) **aturada i adhesió** (*attachment*) i (4) **migració transendotelial** (*migration*) (figura 1).

En el primer pas, els leucòcits s'uneixen dèbilment a l'endoteli mitjançant una interacció de ràpida disociació definida principalment a través de les selectines (durant la resposta inflamatòria, les citocines i altres mediadors actuen sobre l'endoteli local induint l'expressió de molècules d'adhesió de la família de les selectines). Aquestes selectines (ja siguin la L-selectina leucocitària i/o la E- o P-selectines endotelials) s'uneixen de forma dèbil i intermitent amb les glicoproteïnes que reconeixen, provocant la fase d'enganxament o lligament (*tethering*) limfocitari, i tot seguit el rodolament (*rolling*) a



favor del flux sanguini i del gradient d'adhesivitat (haptotaxis) característic d'aquesta arribada limfocitària (figura 1).



**Figura 1:** Els 4 passos del procés d'extravassació leucocitària i les molècules d'adhesió que hi intervenen.

Quan els leucòcits es troben en plena fase de rodolament són activats per diversos estímuls quimioatracients que, o bé es troben permanentment a la superfície de les cèl·lules endotelials, o bé són secretats localment per cèl·lules involucrades en la resposta inflamatòria. Entre aquests quimioatracients destaquen els membres d'una gran família de citocines quimiotàctiques anomenats **quimiocines**. Diverses són les quimiocines involucrades en el procés d'activació [p.e. CXCL8 (IL-8) i CCL4 (MIP-1 $\beta$ )]. Tot i això, no tots els quimioatracients pertanyen al grup de quimiocines. Altres quimioatracients importants són el PAF (*platelet-activating factor*), productes derivats de la cascada proteolítica del complement (com C5a i C3a) i varis pèptids *N*-formil produïts pel trencament de proteïnes bacterianes durant la infecció (figura 1).

La unió d'aquests quimioatracients als receptors presents a la membrana dels neutròfils dispara un senyal activador a través de proteïnes G associades al receptor. Aquest senyal induïx un canvi conformacional en les integrines que es troben a la membrana dels neutròfils, incrementant la seva afinitat per les molècules d'adhesió de la superfamília de les immunoglobulines presents a l'endoteli. La interacció entre aquests dos tipus de CAMs estableix la unió dels neutròfils a les cèl·lules endotelials, permetent-ne una forta adhesió. Seguidament el neutròfil migra a través de la paret del vas cap a l'interior del teixit (figura 1).

### 1.1.5. L'extravassació dels limfòcits

Diverses subpoblacions de limfòcits presenten una extravassació dirigida cap als llocs de lesió (inflamació) i cap als òrgans limfoides (recirculació). La recirculació dels limfòcits està controlada acuradament per assegurar que les poblacions apropiades de cèl·lules B i T siguin reclutades estructuradament cap als diferents teixits. Igual que els neutròfils, l'extravassació dels limfòcits involucra interaccions entre un gran nombre de molècules d'adhesió cel·lular. El procés global és similar al que succeeix durant l'extravassació dels neutròfils (apartat 1.1.4).

Algunes regions de l'endoteli vascular de les vècules postcapilars de diversos òrgans limfoides estan compostes per cèl·lules endotelials especialitzades que presenten una morfologia cuboidal. Aquestes regions s'anomenen **vècules endotelials altes** (HEV, de l'anglès *high endothelial venules*). La característica fonamental de les HEV, que permet l'extravassació limfocitària, és la major adhesivitat de l'endoteli alt (si es compara amb l'endoteli venular de la resta de teixits) pels limfòcits circulants [12]. El desenvolupament i manteniment de les HEV en els òrgans limfoides està influenciat per citocines produïdes en resposta a la captura d'antígens (les HEV no es desenvolupen en animals que creixen

en ambients lliures de patògens), encara que també hi intervenen elements constitutivament expressats.

Les vècules d'endoteli alt expressen una gran diversitat de molècules d'adhesió cel·lular. Com altres cèl·lules d'endoteli vascular, les HEV expressen CAMs de la família de les selectines (E- i P-selectina), de la família de molècules tipus mucina (GlyCAM-1 i CD34) i de la superfamília de les immunoglobulines (ICAM-1, ICAM-2, ICAM-3, VCAM-1 i MAdCAM-1). Algunes d'aquestes molècules d'adhesió estan distribuïdes d'una forma específica de teixit, anomenant-se **adressines vasculares** (VAs) degut a que serveixen per dirigir l'extravassació de diferents poblacions de limfòcits recirculants a determinats òrgans limfoides. Actualment el terme adressina s'aplica en general a les CAMs presents a l'endoteli.

Com s'ha dit, el procés general d'extravassació en la recirculació dels limfòcits és similar al dels neutròfils en la inflamació. Una característica important del primer d'aquests dos processos és que les diferents subpoblacions de limfòcits migren de forma diferencial cap a diferents teixits. Aquest procés s'anomena **nidificació** (*homing*).

Els diferents patrons de recirculació de les poblacions limfocitàries estan dirigits per receptors que es troben a la superfície dels limfòcits recirculants i que reconeixen adressines vasculares particulars que es troben a les HEV dels diferents òrgans limfoides secundaris (i a l'endoteli dels llocs quan es produeix inflamació). Degut a aquesta funció de dirigir la circulació de diverses poblacions de limfòcits a teixits limfoides (o inflamats) concrets, a aquests receptors se'ls anomena **receptors de nidificació** (*homing receptors*).

## 1.2. LES QUIMIOCINES

Les quimiocines són una superfamília de petits polipèptids (8-10 kDa) estructuralment relacionats, que comparteixen la capacitat d'estimular la motilitat dels leucòcits (**quimioquinesi**) així com també el seu moviment dirigit (**quimiotaxi**). El nom "quimiocines" prové de la contracció de citocines **quimiotàctiques**.

Les quimiocines controlen selectivament l'adhesió, quimiotaxi i activació de molts tipus de poblacions i subpoblacions leucocitàries. Per tant, podem considerar que són els principals reguladors del trànsit leucocitari. Una gran varietat de teixits limfoides i no limfoides poden produir quimiocines, sovint en connexió amb la iniciació o progressió de la inflamació (quimiocines inflamatòries). Tot i això, entre les quimiocines, hi ha

membres que estan involucrats en la regulació del trànsit leucocitari normal (que inclou la recirculació limfocitària entre òrgans i teixits limfoides – quimiocines homeostàtiques). Les quimiocines provoquen la migració dels leucòcits cap a l'interior dels teixits per inducció de l'adherència d'aquestes cèl·lules a l'endoteli vascular. Després d'aquesta migració dins els teixits, els leucòcits són atrets als llocs concrets on es localitzen elevades concentracions de quimiocines.

### 1.2.1. Classificació

Una de les característiques estructurals compartida per la majoria de quimiocines és la presència, en la seva seqüència aminoacídica, de quatre cisteïnes conservades entre les quals s'estableixen dos ponts disulfur que estableixen el plegament de la molècula.

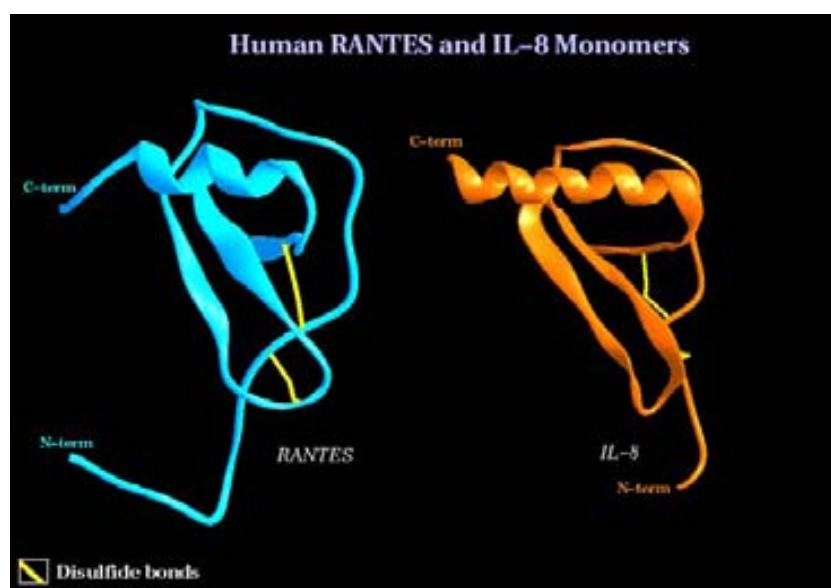
La classificació de les quimiocines es basa en la posició dels dos primers residus cisteïnics de la proteïna madura. Segons aquest criteri es distingeixen quatre subfamílies designades com a CXC ( $\alpha$ ), CC ( $\beta$ ), C ( $\gamma$ ) i CX3C ( $\delta$ ). Totes les quimiocines, excepte tres, pertanyen als subgrups CXC i CC, mentre que els grups C i CX3C només disposen de dos i un membre respectivament [13].

Les quimiocines **CXC** tenen un aminoàcid no conservat separant els dos primers residus de cisteïna de la proteïna, mentre que les **CC** o  $\beta$  quimiocines no presenten cap aminoàcid entre les dues primeres cisteïnes. La limfotactina  $\alpha$  i la  $\beta$  (productes de dos gens altament relacionats) són els únics membres de la subfamília de les **C** quimiocines [14], i tenen la peculiaritat que són les úniques quimiocines que han perdut dues de les 4 cisteïnes que caracteritzen aquesta superfamília i, per tant, disposen d'un únic pont disulfur (tot i això mantenen les característiques funcionals bàsiques que distingeixen la superfamília de les quimiocines). Finalment, la fractalquina, l'únic membre de les quimiocines **CX3C**, ens mostra tres aminoàcids entre les primeres dues cisteïnes. Presenta una característica única entre les quimiocines i és que s'expressa anclada a la membrana cel·lular (tot i que posteriorment es pot detectar una forma soluble i funcional) [15].

Dins d'aquesta classificació bàsica, cal destacar algunes peculiaritats: En un subgrup de CC quimiocines trobem dues cisteïnes addicionals a les quatre típicament conservades. Aquestes quimiocines són la CCL1 (I-309), CCL15 (Leucotactina-1 o HCC-2), CCL23 (MPIF1) i CCL21 (SLC) i poden formar un tercer pont disulfur aprofitant aquestes dues cisteïnes addicionals [16-19]. Una altra quimiocina, la CCL24 (Eotaxina-2), presenta una cinquena cisteïna no aparellada a l'extrem C-terminal [20].

### 1.2.2. Estructura i relacions estructura-funció de les quimiocines

Les dades estructurals aportades en la classificació de les quimiocines són un clar reflex de la importància de l'estructura en la funcionalitat de les quimiocines. És doncs central entendre els elements estructurals bàsics d'aquesta superfamília. Bàsicament es distingeixen 3 dominis estructurals: a) La zona N-terminal, anterior a la primera cisteïna, que no presenta una estructura ordenada (amb l'excepció de CCL5, RANTES, que té una petita regió en forma de làmina- $\beta$  abans d'aquesta primera cisteïna [21]), b) la zona central, amb 3 làmines- $\beta$  antiparal·leles unides per nances no plegades i c) l'extrem C-terminal, plegat com a hèlix- $\alpha$  hidrofòbica. De fet, la triple cadena de làmines- $\beta$  antiparal·leles forma una base sobre la que es disposa la hèlix- $\alpha$  hidrofòbica C-terminal. La relació espacial entre l'N-terminal, les làmines- $\beta$  i l'hèlix- $\alpha$  es manté pels dos ponts disulfur que defineixen la compactació globular de la molècula (figura 2).



**Figura 2:** Estructura tridimensional de dues quimiocines representatives dels subgrups CC (CCL5 o RANTES) i CXC (CXCL8 o IL-8)

Aquests tres dominis estructurals tenen diferents funcions. L'extrem N-terminal és essencial per a la unió als receptors cel·lulars de membrana [22]. A més el propi extrem N-terminal juntament amb la zona de les làmines- $\beta$  proporcionen una conformació adequada per a la formació de les interaccions que defineixen les estructures quaternàries com els dímers de quimiocines. D'altra banda, l'hèlix- $\alpha$  C-terminal d'algunes quimiocines interacciona amb baixa afinitat ( $K_d=10^{-5}M$ ) amb els glucosaminoglicans

(GAGs) presents a la superfície cel·lular i a la matriu extracel·lular [23], encara que solen ser residus de les nances els elements més importants per aquesta unió. Tot i que la capacitat de les quimiocines d'unir-se a GAGs no és essencial per la seva funció, s'ha vist que optimitza la interacció amb el receptor, el manteniment del gradient i la migració quimiotàctica de les cèl·lules que expressen els receptors.

Entre les modificacions post-traduccionals cal destacar que, en general, les quimiocines contenen llocs de glicosilació i, per tant, quan són produïdes de forma natural estan altament glicosilades. Això explica que a les quimiocines recombinants produïdes en bacteris els manquin les glicosilacions, encara que mantenen la capacitat d'unir-se i activar el receptor induint la senyalització i la quimiotaxi cel·lular (*in vitro* i *in vivo*). Per tant, la funció de la glicosilació en les quimiocines encara s'ha de determinar amb exactitud, tot i que es suposa que tindria un paper en la prolongació de la seva vida mitja, alhora que milloraria llur unió a GAGs.

Els diversos estudis de cristal·lització que s'han realitzat permeten definir altres generalitzacions sobre l'estructura de les quimiocines. La conformació d'aquestes proteïnes en forma monomèrica és globular i l'espai existent entre les dues primeres cisteïnes (que és el criteri en que es basa la classificació d'aquesta superfamília) defineix l'orientació que pren el *loop* format per l'extrem N-terminal i la primera làmina  $\beta$ . Aquesta orientació és característica en cadascuna de les 4 famílies de quimiocines i defineix un plegament tridimensional específic per la unió al receptor. Les quimiocines formen dímers a concentracions micromolars i en condicions de pH baix i concentracions de sal elevades [24]. De totes maneres en condicions fisiològiques també es formen dímers que coexisteixen amb els monòmers en un equilibri, desplaçat sempre a favor de les formes monomèriques [25, 26].

Les quimiocines són produïdes com a polipèptids i han de ser tallades adequadament per adquirir la forma funcional secretada. Es necessita una proteolisi N-terminal per convertir el polipèptid en la proteïna madura i activa, tot i que una segona proteolisi N-terminal provoca la seva inactivació. La quimiocina madura i secretada també pot ser tallada extracel·lularment per diversos enzims. Un dels que més destaquen en aquesta funció és el CD26, una dipeptidilpeptidasa altament expressada en cèl·lules T activades (funciona com a molècula coestimulatòria) [27-29]. La glucoproteïna CD26 talla dipèptids de l'extrem N-terminal de proteïnes amb un residu de prolina o alanina en la penúltima posició. En base a aquest criteri, aquesta peptidasa és capaç de tallar els dos primers aminoàcids de l'extrem N-terminal de diverses quimiocines com la CCL5 (RANTES), CXCL6 (GCP-2), CXCL9 (MIG), CXCL10 (IP-10), CXCL11(I-TAC), CXCL12 (SDF), CCL11

(Eotaxina), CCL22 (MDC), CCL3 (MIP-1 $\alpha$ ) i CCL4 (MIP-1 $\beta$ ). Les conseqüències funcionals d'aquesta proteolisi en les diferents quimiocines esmentades són variades (ja que poden afectar a diversos paràmetres) i no segueixen un patró concret [30-32] encara que en alguns casos (com els de CCL3 i CCL4) generen formes de major afinitat al receptor [33, 34].

### 1.2.3. Classificació funcional de les quimiocines CXC

Tot i que cada quimiocina presenta unes peculiaritats funcionals individuals, es poden fer una sèrie de generalitzacions basades en les relacions estructurals dels diferents membres d'aquesta família (a la taula 1 es mostren les propietats de les quimiocines CXC i s'inclouen també els únics membres de les famílies C i CX3C).

#### 1.2.3.1. Funció inflamatòria de les quimiocines CXC (ELR<sup>+</sup>)

Algunes de les quimiocines CXC contenen un motiu tripeptídic (ELR) característic format per àcid glutàmic (E), leucina (L) i arginina (R) a les posicions N-terminals 4, 5 i 6. Aquestes quimiocines són la CXCL8 (IL-8), CXCL7 (NAP-2), CXCL5 (ENA-78), CXCL1 (GRO $\alpha$ ), CXCL2 (GRO $\beta$ ), CXCL3 (GRO $\gamma$ ) i CXCL6 (GCP-2) i comparteixen el fet d'unir-se al receptor CXCR2. Precisament aquest motiu ELR és essencial per la unió d'alta afinitat a CXCR2 [35, 36].

Les quimiocines ELR<sup>+</sup> són produïdes per una gran varietat de cèl·lules en resposta a diversos estímuls (bàsicament citocines proinflamatòries com la IL-1 i el TNF $\alpha$ ). L'expressió d'aquestes quimiocines als llocs d'inflamació i la distribució dels receptors CXCR2 i CXCR1 en neutròfils explica la participació d'aquestes quimiocines en les respostes inflamatòries agudes, en les que els polimorfonuclears tenen un paper protagonista. Així doncs, les quimiocines CXC ELR<sup>+</sup> promouen l'adhesió dels neutròfils a les cèl·lules endotelials, la seva extravassació i la migració al llarg del gradient de quimiocines associades a la matriu extracel·lular fins al lloc d'inflamació. Cal remarcar que en situacions on es troben en concentracions elevades, aquestes quimiocines provoquen l'activació de la capacitat fagocítica dels leucòcits polimorfonuclears, induint fins i tot la seva degranulació.

Quimiocina	Cr.	Cèl·lules Productores	Estímuls Inductors	Recep.	Efectes
CXCL1,2,3 (GRO $\alpha,\beta,\gamma$ / MGS $\alpha,\beta,\gamma$ )	4q13-21	Monòcits, fibroblasts, cèl·lules endotelials	IL-1, LPS, TNF	CXCR2	Reclutament i activació de neutròfils Angiogènesi
CXCL4 (PF4)	4q12-21	Plaquetes	Agregació plaquetària	no definit	Angioestàtica Procoagulant
CXCL5 (ENA-78)	4q13-21	Cèl·lules endotelials, fibroblasts, monòcits, queratinòcits	IL-1, LPS, TNF $\alpha$	CXCR2	Reclutament i activació de neutròfils Angiogènesi
CXCL6 (GCP-2)	4q12-13	Cèl·lules d'osteosarcoma	IL-1 $\beta$ , LPS	CXCR2 CXCR1	Reclutament i activació de neutròfils
CXCL7 (NAP-2)	4q12-13	Plaquetes, monòcits	LPS	CXCR2 CXCR1	Reclutament i activació de neutròfils
CXCL8 (IL-8)	4q12-21	Monòcits, neutròfils, cèl·lules T, cèl·lules NK, cèl·lules endotelials i epitelials	LPS, IL-1, TNF, IL-3, IL-7, IL-13, H <sub>2</sub> O <sub>2</sub> , hipòxia	CXCR2 CXCR1	Reclutament i activació de neutròfils Angiogènesi Modula l'hematopoiesi
CXCL9 (MIG)	4q21	Monòcits, cèl·lules hepàtiques	IFN $\gamma$	CXCR3	Atrau cèl·lules T activades. Angioestàtica
CXCL10 (IP-10)	4q21	Monòcits, cèl·lules T i endotelials, queratinòcits	IFN $\alpha,\beta,\gamma$ TNF $\alpha$ , LPS	CXCR3	Atrau cèl·lules T activades. Angioestàtica
CXCL11 (I-TAC)	4q21	Monòcits, astròcits	IFN $\beta$ , IL-1	CXCR3	Atrau cèl·lules T activades. Angioestàtica
CXCL12 (SDF-1)	10q11	Àmpliament expressada (cèl·lules estromals de medul·la òssia)	Constitutiva	CXCR4	Atrau cèl·lules B, T i CD34 <sup>+</sup> Desenvolupament de varis teixits Angiogènesi
CXCL13 (BLC / BCA-1)	4q21	Ganglis limfàtics, melsa, fetge	Constitutiva	CXCR5	<i>Homing</i> de cèl·lules B cap a òrgans limfoides secundaris
XCL1 / XCL2 (Limfotactina $\alpha/\beta$ ) (subfamília C)	1q23	Cèl·lules T CD8, Mastòcits, NK, NKT i $\gamma\delta$ epidèrmiques	Mitògens	XCR1	Atrau cèl·lules T i cèl·lules NK
CX3CL1 (Fractalquina) (subfamíl. CX3C)	16q13	Cèl·lules endotelials, monòcits, microglia	TNF $\alpha$ , IL-1	CX3CR1	Atrau monòcits, cèl·lules T i NK Reparació neuronal

**Taula 1:** Característiques principals de les quimiocines de les famílies CXC, C i CX3C

### 1.2.3.2. Altres funcions vasculars de les quimiocines CXC

Les quimiocines CXC actuen sobre els vasos amb efectes angiogènics i angioestàtics. Així, les quimiocines CXC ERL<sup>+</sup> tenen un marcat efecte angiogènic. Un exemple d'això el trobem en el fet que si els tumors que expressen constitutivament CXCL8 (IL-8), CXCL1,2,3 (GRO $\alpha,\beta,\gamma$ ) o CXCL5 (ENA-78) són tractats amb anticossos neutralitzants contra aquestes quimiocines, mostren un creixement reduït i una menor capacitat de metastasi que es basa en l'absència de neovascularització [37]. De forma oposada,



determinades quimiocines CXC sense el motiu ELR (ELR<sup>-</sup>) són angioestàtiques (disminueixen la producció vascular): CXCL4 (PF4), CXCL9 (MIG), CXCL10 (IP-10) i CXCL11 (I-TAC). L'efecte angioestàtic de CXCL4 ha estat clínicament avaluat com un possible tractament antitumoral en humans [38, 39] i les altres tres quimiocines, que comparteixen la unió al receptor CXCR3, també han demostrat un efecte antitumoral basat en la seva interferència amb l'angiogènesi [37, 40].

Els interferons  $\alpha$ ,  $\beta$  i  $\gamma$ , coneguts pels seus efectes angiostàtics, no tan sols suprimeixen la producció de les quimiocines angiogèniques CXC ELR<sup>+</sup> si no que també estimulen la producció de quimiocines angioestàtiques CXC ELR<sup>-</sup>. És a dir, els efectes angioestàtics dels interferons poden estar mitjançats pel canvi de producció de quimiocines CXC angiogèniques a quimiocines CXC angioestàtiques.

Tot i que les principals quimiocines CXC angiogèniques són ELR<sup>+</sup>, en trobem d'ELR<sup>-</sup> amb aquests mateixos efectes; és el cas de les dues variants de CXCL12 (SDF-1 $\alpha$  i SDF-1 $\beta$ ), que són clares inductores de la proliferació vascular [41].

### **1.2.3.3. Funció homeostàtica de les quimiocines CXC**

Diverses dades fan que la quimiocina CXCL12 (SDF-1) sigui considerada com la quimiocina més primitiva. Per començar és la única que es localitza al cromosoma 10, separada de totes les altres quimiocines CXC, que es troben al cromosoma 4. De fet, CXCL12 es comporta més com una hormona hematopoiètica homeostàtica (crítica pel *homing* i l'angiogènesi) que no pas com una quimiocina proinflamatòria (nivells baixos de CXCL12 es troben constitutivament al plasma) [42]. Potser la dada més suggerent alhora de considerar CXCL12 com la quimiocina més primitiva és que exhibeix una homologia de seqüència equidistant a les CXC i CC quimiocines i, al contrari d'altres quimiocines que són força divergents entre espècies, CXCL12 està altament conservada en l'evolució (per exemple, només té un únic aminoàcid diferent entre la forma humana i murina) [43]. Aquesta quimiocina està expressada per un ampli ventall de cèl·lules i teixits i té efectes quimiotàctics també sobre una gran varietat de cèl·lules, degut a l'extensa distribució del seu receptor (CXCR4) [44]. En aquest sentit, s'ha demostrat clarament que CXCL12 és essencial per la vida: ratolins amb absència d'aquesta quimiocina (o del seu receptor), per eliminació del gen mitjançant recombinació homòloga, moren poc després del naixement [41, 45], fet que no s'ha observat, de moment, en cap dels *knockouts* realitzats per a altres quimiocines. Cal destacar una última característica de CXCL12 i és la seva capacitat d'inhibir l'entrada de soques X4 de l'HIV en línies cel·lulars que expressen el receptor CXCR4 [46, 47].

Una altra quimiocina CXC, la CXCL13 (BLC o BCA-1), únic lligand conegut pel receptor CXCR5, presenta únicament funcions no inflamatòries. Està constitutivament expressada a alts nivells als ganglis limfàtics, melsa, fetge i als fol·licles limfoides de les plaques de Peyer. La seves funcions principals inclouen el *homing* de cèl·lules B i el desenvolupament normal del teixit limfoide [48-50]. Per tant, CXCL13 és més representativa d'una classe de citoquines reguladores homeostàtiques dels òrgans limfoides secundaris que no pas un membre pertanyent a les quimiocines CXC proinflamatòries.

#### 1.2.4. Funcions de les quimiocines CC

Les quimiocines CC actuen, en general, sobre cèl·lules mononuclears més que no pas sobre neutròfils (diana principal de les quimiocines CXC). Aquesta subfamília de quimiocines la podem dividir en 2 grups: les que presenten activitat proinflamatòria i les que tenen un paper en el desenvolupament i l'homeòstasi.

##### 1.2.4.1. Grup de quimiocines CC proinflamatòries

Dins aquest grup de quimiocines proinflamatòries podem distingir tres subgrups: 1-subgrup MCP/eotaxina, 2-subgrup RANTES/MIP i 3-subgrup HCC.

Els membres del subgrup MCP/eotaxina estan altament relacionats tant en la localització cromosòmica com en l'estructura proteica i la utilització dels receptors [51]. Aquest subgrup és d'especial interès en els fenòmens d'hipersensibilitat de tipus I (al·lèrgics) ja que entre els seus components trobem alguns potents quimioattractors d'eosinòfils i diversos que poden mitjançar l'alliberament d'histamina [52].

La CCL2 (MCP-1) va ser inicialment descoberta com un producte de monòcits induït per la IL-1 amb una potent activitat quimioattractora sobre monòcits [53, 54]. CCL2 comparteix un 65% de la seqüència aminoacídica amb CCL8 (MCP-2), CCL7 (MCP-3) i CCL13 (MCP-4) i, mentre totes elles utilitzen el receptor CCR2, CCL8, CCL7 i CCL13 també utilitzen CCR1 i CCR3 i, a més a més, CCL8 també utilitza CCR5. Aquestes 4 quimiocines atrauen basòfils i mastòcits i poden induir ràpidament la degranulació dels basòfils amb la conseqüent alliberació d'histamina [52].

La CCL11 (Eotaxina-1) utilitza únicament el receptor CCR3 i, igual com la CCL13, atrau fortament eosinòfils i basòfils [55].

L'últim membre descobert d'aquest subgrup és la CCL24 (Eotaxina-2 o MPIF2), que també forma part de les quimiocines al·lèrgèniques ja que s'uneix al receptor CCR3 [20].

Tot i aquesta contribució a les reaccions al·lèrgiques, moltes vegades amb efectes solapats entre ells, els membres d'aquest subgrup també mostren altres activitats proinflamatòries. Per exemple, les quimiocines CCL2, 8, 7 i 13 (MCP-1,-2,-3,-4) són quimioatracients per monòcits i cèl·lules T i participen en les reaccions inflamatòries cròniques.

Hi ha una quimiocina, la CCL1 (I-309), que tot i estar localitzada prop del subgrup de quimiocines al·lèrgèniques en el dendograma de divergència genètica, no se la considera pertanyent a aquest subgrup. No obstant això, pot participar en respostes IgE-dependents ja que utilitza el receptor CCR8, que s'expressa preferencialment en cèl·lules T CD4<sup>+</sup> productores de citocines Th2 [56-58].

Un segon subgrup de quimiocines CC proinflamatòries homòlogues inclou CCL5 (RANTES), CCL3 (MIP-1 $\alpha$ ), CCL4 (MIP-1 $\beta$ ) i CCL18 (PARC). Aquesta última quimiocina (CCL18 o PARC), tot i que està relacionada a nivell de seqüència amb altres quimiocines proinflamatòries, encara no s'ha pogut relacionar a nivell funcional amb aquest subgrup ja que el seu receptor no s'ha descobert [59].

Pel que fa als altres membres, CCL3, CCL4 i CCL5 comparteixen la capacitat quimioatracient per a monòcits i diferents subgrups de limfòcits T. CCL5 atrau basòfils i indueix l'alliberament d'histamina. A més a més, també exerceix quimioatracció sobre eosinòfils, igual que CCL3 en certes ocasions. Tot i això, aquestes dues quimiocines no es consideren contribuïdores principals de les respostes al·lèrgiques, ja que l'aplicació d'anticossos neutralitzants contra elles no redueix significativament aquestes respostes [52]. CCL5 participa en respostes inflamatòries cròniques immunològicament regulades.

CCL3 i CCL4 són dues quimiocines estructuralment i funcional relacionades però cadascuna exhibeix diferents característiques que els permeten regular independentment aspectes específics de la resposta inflamatòria. L'aspecte més ben caracteritzat de CCL3 i CCL4 és la seva capacitat de regular el trànsit [60] i la capacitat d'activació de determinats subgrups de cèl·lules inflamatòries [61-63]. Un lligam important entre aquestes dues quimiocines i CCL5 és el fet que el receptor CCR5 (utilitzat per les tres quimiocines) funciona com un co-receptor de les soques R5 de l'HIV necessari per la seva entrada a les cèl·lules (s'han demostrat efectes antivirals de CCL3,4,5 com a resultat de la seva interferència amb la utilització del CCR5 per part de l'HIV) [64, 65]. En qualsevol

cas, donat que CCL4 és l'objecte d'aquest estudi, presentem més informació a l'apartat 1.4.

El tercer i últim subgrup de quimiocines CC amb funcions proinflamatòries inclou CCL14 (HCC-1), CCL15 (HCC-2), CCL16 (HCC-4) i CCL23 (CK $\beta$ 8-1 / MPIF1). Tot i que no és molt habitual entre les quimiocines CC, CCL14 s'expressa constitutivament en molts teixits sense activació i està present al plasma humà a concentracions nanomolars [17]. Encara que utilitzi el receptor CCR1, la seva acció quimioatracent es limita als monòcits i és 100 vegades menor en comparació a CCL3 (que també utilitza aquest mateix receptor).

CCL15 mostra una expressió molt més restringida [66] i utilitza el receptor CCR1 de forma similar a CCL3 (amb efectes potents sobre monòcits i limfòcits, i modestos sobre eosinòfils).

Finalment, CCL23 és una quimiocina molt similar en seqüència a CCL15. De fet, excepte per la seva incapacitat d'atraure eosinòfils, CCL23 sembla indistingible de CCL15.

CCL16 presenta una característica inusual ja que la seva expressió per part de monòcits augmenta per la IL-10, una citocina que disminueix l'expressió de moltes altres quimiocines i citocines [67].

#### **1.2.4.2. Grup de quimiocines CC homeostàtiques i del desenvolupament**

Hi ha una sèrie de quimiocines CC que tenen una funció més important en el desenvolupament, trànsit homeostàtic i *homing* de diversos subgrups limfoides que no pas en la inflamació. Són la CCL25 (TECK), CCL22 (MDC), CCL17 (TARC), CCL21 (SLC o 6CKine), CCL19 (ELC o MIP-3 $\beta$ ) i CCL20 (LARC). CCL25 (TECK) està altament expressada en el timus adult i està involucrada en el trànsit limfoide cap al timus i en el seu interior (s'ha proposat que CCL25 pot induir la migració de timòcits des de l'escorça tímica a la medul·la i pot ser important en la repoblació del timus post-irradiació). CCL22 (MDC) i CCL17 (TARC) comparteixen un 37% d'identitat aminoacídica i la utilització del receptor CCR4, i ambdues s'expressen de forma elevada en el timus (s'ha proposat que CCL17 té un paper de guia en la migració de timòcits CD4<sup>+</sup> positivament seleccionats des de l'escorça a la medul·la per a la seva selecció negativa). Finalment les quimiocines homeostàtiques CCL21 (SLC), CCL19 (ELC) i CCL20 (LARC) són peculiars en el fet que es produeixen constitutivament en teixits limfoides i també perquè estan localitzades en el cromosoma 9 (CCL21 i CCL19) o 2 (CCL20) en comptes del 17, que és on es troben els altres gens de quimiocines CC proinflamatòries.

Els teixits i tipus cel·lulars que expressen aquestes quimiocines i les seves funcions principals, així com les de la resta de quimiocines CC es resumeixen a la taula 2.

Nom	Cr.	Cèl·lules Productores	Estímul Inductors	Recep.	Efectes
CCL1 (I-309)	17q11	Cèls. T activades, monòcits	IgG, LPS	CCR8	Atrau monòcits i limfòcits Th2
CCL2 (MCP-1)	17q11	Fibroblasts, cèls. epitelials i endotelials, eosinòfils, monòcits, astròcits	IFN $\gamma$ , IL-1, TNF, IL-6, IL-13, IL-15	CCR2	Atrau monòcits, limf. T memòria activats i basòfils (histamina) Polarització Th2
CCL3 (MIP-1 $\alpha$ )	17q11	Monòcits, cèls. T i B, microglia, cèls. dendrítiques i NK, fibroblasts	LPS, IL-1 $\beta$ , TNF $\alpha$ , IL-2, IL-6	CCR1 CCR4 CCR5	Atrau neutròfils, monòcits i limfòcits CD8 <sup>+</sup> Polarització Th1 Inhibeix l'hematopoiesis
CCL4 (MIP-1 $\beta$ )	17q11	Monòcits, cèls. T i B, microglia, cèls. dendrítiques i NK, fibroblasts	LPS, IL-1 $\beta$ , TNF $\alpha$	CCR5 CCR8	Atrau neutròfils, monòcits i limfòcits T CD4 <sup>+</sup> Polarització Th1
CCL5 (RANTES)	17q11	Fibroblasts, cèls. epitelials i endotelials, monòcits, cèls. T	IL-1 $\beta$ , TNF $\alpha$ , Angiotensina II	CCR1 CCR3 CCR5	Atracció i activació de basòfils, eosinòfils i limfòcits T (també monòcits i dendrítiques)
CCL7 (MCP-3)	17q11	Monòcits, cèl·lules endotelials, fibroblasts, astròcits	IL-4, IL-13	CCR2 CCR1 CCR3	Atrau monòcits, limf. T memòria activats i basòfils (histamina)
CCL8 (MCP-2)	17q11	Fibroblasts, neutròfils, cor, pulmó, timus, pàncreas, intestí	IL-1 $\beta$ , IFN $\gamma$	CCR2 CCR3 CCR1 CCR5	Atrau monòcits, limf. T memòria activats, basòfils (histamina) i mastòcits
CCL11 (Eotaxina)	17q11	Cèls. epitelials i endotelials, monòcits	IL-4, TNF $\alpha$	CCR3	Atracció d'eosinòfils (acció principal) basòfils i cèls. Th2
CCL13 (MCP-4)	17q11	Macròfags, cèl·lules endotelials, intestí prim, colon	IL-1, TNF, IFN $\gamma$	CCR2 CCR3	Atrau eosinòfils, basòfils i mastòcits
CCL14 (HCC-1)	17q11	Àmplia distribució, teixits limfoides	Constitutiva	CCR1	Proliferació de progenitors de medul·la òssia
CCL15 (HCC-2)	17q12	Monòcits, fetge, intestí prim, colon	TNF $\alpha$ , IFN $\gamma$ , LPS, IL-4	CCR1 CCR3	Atrau monòcits i cèls. T Inhibeix la proliferació de progenitors de medul·la òssia
CCL16 (HCC-4)	17q11	Monòcits, fetge	IL-10	CCR1	Atrau monòcits i cèls. T Inhibeix la proliferació de progenitors de medul·la òssia
CCL17 (TARC)	16q13	Timus, pulmó, intestí prim, colon	PHA	CCR4	Atrau limfòcits Th2
CCL18 (PARC)	17q11	Macròfags, cèls. dendrítiques, pulmó, òrgans limfoides	LPS, IL-4, IL-13, IL-10	no definit	Atrau cèl·lules T CD45RA <sup>+</sup>
CCL19 (ELC)	9p13	Timus, ganglis limfàtics, amígdala	Constitutiva	CCR7	Indueix adhesió i <i>homing</i> de cèl·lules T verges a les àrees parafol·liculars
CCL20 (LARC)	2q35	Teixits limfoides, fetge, timus, amígdala, monòcits i DC	Constitutiva	CCR6	Inhibeix la proliferació de progenitors de m. òssia Tràfic de cèls. T i DC
CCL21 (SLC)	9p13	Ganglis limfàtics, melsa, tiroides, timus, monòcits i DC	Constitutiva augmenta per la IL-10	CCR7	<i>Homing</i> de cèl·lules T i cèl·lules dendrítiques en teixits limfoides

CCL22 (MDC)	16q13	Ganglis limfàtics, timus, macròfags, DC	Constitutiva	CCR4 CCR8	Atrau cèl·lules NK, DC i cèl·lules Th2
CCL23 (MPIF1)	17q11	Pulmó, fetge, medul·la òssia, placenta	IFN $\beta$ i IFN $\gamma$	CCR1	Inhibeix la proliferació de progenitors de m. òssia
CCL24 (eotaxina-2)	7q11	Cèl·lules T activades, macròfags	GM-CSF anti-CD3 / IL-2	CCR3	Atrau eosinòfils i basòfils Inhibeix la proliferació de progenitors de m. òssia
CCL25 (TECK)	19p13	Timus (cèl·lules dendrítiques i estromals tímiques)	LPS	CCR9 CCR10	Atrau monòcits, DC tímocits. Desenvolupament de cèls. T al timus

**Taula 2:** Característiques principals de les quimiocines de la família CC

### 1.3. ELS RECEPTORS DE QUIMIOCINES

Els receptors de quimiocines es defineixen per la seva habilitat per unir quimiocines d'una forma específica i saturable i, seguidament, transduir una resposta cel·lular. A nivell molecular, comparteixen el fet de ser receptors amb set dominis transmembrana acoblats a proteïnes G [68-72].

Les seqüències aminoacídiques dels diferents receptors de quimiocines (que contenen de 340 a 370 aminoàcids) tenen d'un 25% a un 80% d'identitat i està acceptat que deriven d'un ancestre molecular comú.

La unió de les quimiocines al seu receptor activa moltes vies de transducció de senyals, la qual cosa resulta en una gran varietat de modificacions en la fisiologia de la cèl·lula diana [73, 74], moltes d'elles associades a canvis en el citoesquelet i les molècules associades a ell.

Com ja s'ha vist anteriorment, aquests receptors poden unir múltiples quimiocines i viceversa. Pel que fa a la seva expressió en diferents tipus cel·lulars, també ens trobem que una cèl·lula determinada pot expressar receptors diversos i que cèl·lules diferents poden expressar un mateix receptor. La taula 3 mostra els receptors principals que expressen els diferents tipus cel·lulars.

La principal funció compartida per les quimiocines i els seus receptors és la quimiotaxi de leucòcits. Aquesta propietat juntament amb l'expressió diferencial dels receptors en determinats tipus cel·lulars fa que les quimiocines i els seus receptors siguin capaços d'organitzar el trànsit de leucòcits *in vivo* d'una forma altament específica. A més a més l'existència d'una certa redundància en aquesta organització permet una regulació molt fina d'aquests processos.

Nom	Neutròfils	Cèl·lules T Resting	Cèl·lules T activades	Subpoblació cèl·lules T	Cèl·lules NK	Cèl·lules B
CXCR1	X	X				
CXCR2	X	X				
CXCR3			X	Th1>>Th2	X	X
CXCR4	X	X	X			X
CXCR5			X	RO+, IL-2R $\alpha$		X
CX3CR1	X		X	CD8	X	
CCR1			X		X	
CCR2			X			
CCR3			X	Th2	X	
CCR4			X	Th2, RO+	X	
CCR5			X	Th1>Th2	X	
CCR6		X		CD4		X
CCR7		X	X	Th1, CD4, CD8		X
CCR8			X	Th2		X
XCR1			X		X	

Nom	Monòcits	Cèl·lules dendrítiques immadures	Cèl·lules dendr. madures	Basòfils	Eosinòfils
CXCR1		X			
CXCR2					
CXCR3	X				X
CXCR4	X		X		
CXCR5					
CX3CR1	X				
CCR1	X	X			X
CCR2	X	X	X	X	
CCR3	X			X	X
CCR4					
CCR5	X	X			
CCR6		X			
CCR7			X		
CCR8	X	X			
XCR1					

**Taula 3:** Receptors de quimiocines i el seu patró d'expressió principal en les diferents subpoblacions leucocitàries.

A part del paper dels receptors de quimiocines en el desenvolupament del sistema immunitari, la defensa de l'hoste i la inflamació, alguns d'ells poden actuar circumstancialment com a factors pro-microbians resultat de l'explotació que en fan certs microorganismes [75, 76]. A part del cas de l'HIV (ja comentat anteriorment) trobem altres virus que indueixen l'expressió de receptors virals de quimiocines en les cèl·lules que infecten, com un possible mecanisme d'evasió del sistema immunitari per segrest de quimiocines (en són exemples el citomegalovirus i l'herpesvirus VIII) [77-79].

#### 1.4. LA QUIMIOCINA CCL4 (MIP-1 $\beta$ )

L'element central d'aquesta tesi, la quimiocina CCL4 (MIP-1 $\beta$ ), és un membre de la família de quimiocines CC altament relacionat amb CCL3 (MIP-1 $\alpha$ ) que regula aspectes de la resposta inflamatòria. CCL4 s'inclou clarament dins el subgrup de quimiocines induïbles que tenen un paper fonamental en la regulació de la resposta de l'hoste envers la invasió de bacteris, virus, paràsits i fongs.

##### 1.4.1. Noms alternatius

El nom comú més conegut de CCL4 és MIP-1 $\beta$  (*macrophage inflammatory protein-1 $\beta$* ), producte del gen *SCYA4*. Altres designacions que ha rebut són ACT-2 (*T cell activation protein 2*), LAG-1 (*lymphocyte activation gene-1*), pAT744, hH400, SISg, HC21, G-26 i MAD-5.

##### 1.4.2. Descripció estructural de la proteïna

CCL4 es sintetitza com un precursor de 92 aminoàcids (10.211 Da) que s'ha de tallar per generar la proteïna madura, de 69 aminoàcids (7818 Da). Els dos ponts disulfur es formen entre els residus Cys11-Cys35 i Cys12-Cys51 [80].

Aquesta quimiocina, de forma similar a CCL3, es presenta en forma de dímers o tetràmers en solució (fins i tot quan es troba en concentracions molt altes i en condicions de baixa salinitat pot formar agregats d'elevat pes molecular) [24, 81, 82]. Una altra característica important és que CCL4 s'uneix a l'heparina amb una alta afinitat [83]. Això és degut a un motiu d'unió a glicosaminoglicans que s'ha identificat per mutagènesi dirigida i que implica varis aminoàcids bàsics: R18, K45 i, sobretot, R46 [84]. Aquesta propietat facilita la seva interacció amb GAGs presents a la superfície cel·lular, incrementant la concentració de quimiocina local i facilitant la seva oligomerització.



### 1.4.3. Principals cèl·lules i teixits que expressen CCL4

En general, les diferents cèl·lules i teixits no expressen constitutivament CCL4 sinó que aquesta expressió s'indueix com a conseqüència d'una activació cel·lular ja sigui per una estimulació per citocines, per certes infeccions intracel·lulars o per determinants específics de patologies concretes [85, 86]. Tenint en compte aquest fet, CCL4 s'indueix en els següents tipus cel·lulars: monòcits / macròfags, cèl·lules dendrítiques, limfòcits T i B, cèl·lules NK, neutròfils, mastòcits, queratinòcits, fibroblasts, cèl·lules de la microglia i astròcits.

### 1.4.4. Receptors

CCL4 s'uneix principalment a CCR5 i, molt secundàriament, a CCR8 (encara que la rellevància fisiològica de la unió a aquest receptor encara no està molt clara) [87-90]. Per tant, és bàsicament a través de CCR5 que CCL4 desenvolupa les seves funcions. CCR5 s'expressa principalment en monòcits/macròfags, cèl·lules dendrítiques, cèl·lules T memòria i cèl·lules de la microglia i, com altres receptors de quimiocines inflamatòries, està involucrat en el reclutament d'aquestes poblacions cel·lulars en llocs d'inflamació. CCR5 va ser descrit inicialment com a receptor per a CCL3, CCL4 i CCL5 [91] i, posteriorment, CCL8 i CCL13 també s'hi van afegir [92-94]. Les citocines proinflamatòries (TNF $\alpha$  i IL-12) i les citocines Th1 (IFN $\gamma$  i IL-2) augmenten l'expressió de CCR5 en la superfície dels PBMCs [95, 96]. S'ha demostrat que CCR5 constitueix el principal correceptor de les soques M-tròpiques dels lentivirus HIV-1, HIV-2 i SIV. L'existència d'un al·lel no-funcional del receptor (CCR5 $\Delta$ 32) que proporciona una forta protecció en front a la infecció pel virus HIV, demostra el paper central de CCR5 en aquesta malaltia [97].

### 1.4.5. Funcions principals

Pel què fa al trànsit i a la mobilització cel·lular, CCL4 provoca un reclutament de monòcits i de subpoblacions específiques de limfòcits T i B al lloc d'inflamació [85, 86, 98]. CCL4 atrau preferentment limfòcits T CD4<sup>+</sup> (mentre que CCL3 ho fa preferentment sobre limfòcits T CD8<sup>+</sup>) i també exerceix quimioatracció per a cèl·lules dendrítiques [85, 86]. Un aspecte que cal destacar és que CCL4 és la quimiocina que exerceix una atracció més potent en cèl·lules T reguladores CD4<sup>+</sup>CD25<sup>+</sup>. El treball on es va descriure aquest efecte [99], compara el perfil de subpoblacions atretes per les diferents quimiocines que secreten les cèl·lules presentadores d'antigen professionals activades. De les diferents quimiocines analitzades (XCL1, CCL2, CCL3, CCL4, CCL5 i CCL22), CCL4 va mostrar una

atracció específica de cèl·lules T reguladores CD4<sup>+</sup>CD25<sup>+</sup>, mentre que en la resta de quimiocines, aquest efecte era residual. A més a més, recentment, s'ha descrit una subpoblació de cèl·lules T reguladores CD8<sup>+</sup>, LAG3<sup>+</sup>, CD25<sup>+</sup>, FOXP3<sup>+</sup>, que exerceixen un efecte supressor sobre cèl·lules T a través de la secreció de CCL4, la qual inhibeix l'activació d'aquestes cèl·lules T interferint en la senyalització del TCR [100].

En l'hematopoiesi CCL4, al contrari que CCL3, no exhibeix una activitat supressora del creixement de progenitors immadurs de medul·la òssia (fins i tot és capaç de bloquejar l'acció supressora de CCL3 sobre aquestes cèl·lules) [101].

Finalment, podem destacar un paper immunoregulator de CCL4 a dos nivells: a l'inici de la resposta immunitària, per la seva capacitat d'atraure cèl·lules dendrítiques al lloc de la infecció, i en una fase més avançada per la generació d'una resposta preferencial del subgrup de limfòcits Th1 i cèl·lules T reguladores (degut probablement a la elevada expressió de CCR5 en aquestes cèl·lules) [85, 86, 102].

#### 1.4.6. Intervenció en patologia

CCL4 està implicada en un ampli ventall de situacions inflamatòries agudes o cròniques que inclouen: malalties infeccioses (infecció per HIV, adenovirus, tuberculosi, sepsis causades per bacteries gram-negatives, meningitis bacteriana, toxoplasmosis...), malalties autoimmunitàries (esclerosi múltiple, alzheimer, poliomiiositis, dermatomiiositis, artritis reumatoide, diabetis tipus I...) al·lèrgies i altres situacions inflamatòries variades. De nou cal destacar en aquest apartat la relació de CCL4 amb la infecció causada per l'HIV, ja que el seu receptor (CCR5) és utilitzat per les soques R5 d'aquest virus com a coreceptor per entrar a les cèl·lules diana. CCL4 és capaç d'inhibir l'entrada de l'HIV R5 via ocupació del seu receptor CCR5 [64, 65] i, per tant, és un potent antiviral natural.

#### 1.4.7. Localització cromosòmica i altres aspectes importants

En humans, el gen de CCL4 (*CCL4* o *SCYA4*) es localitza al cromosoma 17q21-23 [103]. En aquesta regió trobem també els gens d'altres quimiocines CC relacionades [16]. En el ratolí, el gen s'anomena també *CCL4* o *SCYA4* i està localitzat al cromosoma 11, en un *cluster* que conté també altres gens de quimiocines CC [104]. Tan en humans com en ratolins, aquesta quimiocina es troba en una configuració de tres exons / dos introns (igual que altres quimiocines CC).

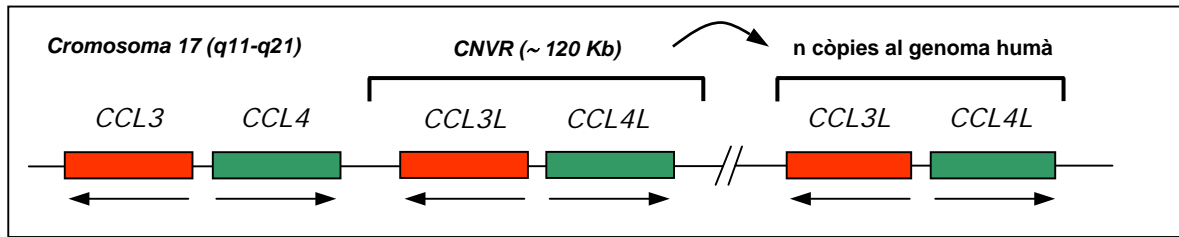
En el cas d'humans, el gens *CCL4* i *CCL3*, presenten una còpia no al·lèlica adicional sorgida possiblement per duplicació gènica i que anomenem *CCL4L* i *CCL3L* (la lletra "L"

que segueix al nom original prové de la paraula anglesa "like"). La homologia d'ambdues parelles de gens (*CCL4-CCL4L* i *CCL3-CCL3L*) és molt elevada, malgrat que existeixen petites diferències a nivell de seqüència aminoacídica [103, 105]. Entre *CCL4* i *CCL4L* la homologia supera el 95% (a nivell nucleotídic). De fet, les proteïnes resultants difereixen en tan sols 3 aminoàcids, dos dels quals estan situats al pèptid senyal i el tercer dins la seqüència de la proteïna madura (homologia >98% a nivell aminoacídic entre les proteïnes madures de *CCL4* i *CCL4L*). Aquest únic canvi és d'una serina (present a *CCL4*) a una glicina (present a *CCL4L*) i està situat a la posició 70 (posició 47 a la proteïna madura) (figura 3). Aquest mateix canvi aminoacídic el trobem també en el cas de *CCL3-CCL3L*. Les conseqüències d'aquest canvi a nivell tridimensional poden ser importants ja que la Serina<sub>70</sub> (*CCL4*) forma un pont d'hidrogen amb l'aminoàcid Thr<sub>67</sub> que estabilitza el gir  $\beta 3$  (o nança 3) que uneix les làmines  $\beta 2$  i  $\beta 3$ . En canvi, l'aminoàcid Gly<sub>70</sub> (*CCL4L*) és incapaç de formar aquest pont d'hidrogen. L'absència d'aquest enllaç causa probablement una cert grau de desestructuració d'aquesta zona, que està implicada en la unió de *CCL4* a GAGs i també en la formació de dímers d'aquesta quimiocina. De totes maneres, fins al moment existeixen molt poques dades sobre diversitat funcional entre *CCL4* i *CCL4L*. En l'únic estudi publicat, Modi et al. [106] no troben diferències significatives entre ambdues quimiocines a nivell de quimiotaxi, afinitat d'unió a CCR5 i bloqueig de la infecció per HIV. Malgrat això, cal tenir en compte que en aquest treball tots els experiments s'han realitzat *in vitro*, sense l'efecte essencial de la unió d'aquestes quimiocines a GAGs.

Recentment aquesta duplicació *CCL4-CCL4L* i *CCL3-CCL3L*, que es va posar de manifest per primer cop al moment del descobriment d'aquestes quimiocines [103], ha resultat tenir una gran transcendència ja que s'ha vist que els locus *CCL4L* i *CCL3L* es troben en nombre variable de còpies en el genoma humà [107] (figura 4). Això significa que existeix variabilitat interindividual en el nombre de còpies d'aquests gens. El rang aproximat de còpies que ens podem trobar del tàndem *CCL4L-CCL3L* és de 0 a 10 per genoma diploide [107]. Per tant, els gens *CCL4L-CCL3L* es troben dins el que s'anomena una regió de nombre variable de còpies (CNVR, de l'anglès *Copy Number Variable Region*). Les primeres evidències de la repercussió d'aquesta variació de nombre de còpies o CNV (de l'anglès *Copy Number Variation*) les trobem en treballs recents on es demostra la relació del nombre de còpies de *CCL3L* amb la susceptibilitat a malalties com l'artritis reumatoide [108], o la infecció pel virus HIV [109].

* 20 * 40 * 60	CAGCACAGGACACAGCTGGGTTCTGAAGCTTCTGAGTTCTGCAGCCTCACCTCTGAGAAA	CCL4
	-----A-----	CCL4L
* 80 * 100 * 120	ACCTCTTTTCCACCAATACCAATGAAGCTCTGCGTGAAGTCTCCTGCTCTCCTCCTAG	CCL4
	-----G-----	CCL4L
	Met <sub>12</sub> →Val	
* 140 * 160 * 180	TAGCTGCCTTCTGCTCTCCAGCGCTCTCAGCACCAATGGGCTCAGACCTCCCACCGCCT	CCL4
	-----T-----A-----	CCL4L
	Pro <sub>20</sub> →Leu Exó-2	
* 200 * 220 * 240	GCTGCTTTTCTTACACCGCGAGGAAGCTTCTCGCAACTTTGTGGTAGATTACTATGAGA	CCL4
	-----	CCL4L
* 260 * 280 * 300	CCAGCAGCCTCTGCTCCCAGCCAGCTGTGGTATTCCAACCAAAAAGAGCAAGCAAGTCT	CCL4
	-----G-----	CCL4L
	Exó-3 Ser <sub>70</sub> →Gly	
* 320 * 340 * 360	GTGCTGATCCCAGTGAATCCTGGGTCCAGGAGTACGTGTATGACCTGGAAGTGAAGTGA	CCL4
	-C-----C-----G-----	CCL4L
* 380 * 400 * 420	CTGCTCAGAGACAGGAAGTCTTTCAGGGAAGGTCACCTGAGCCCGGATGCTTCTCCATGAG	CCL4
	-----T-----	CCL4L
* 440 * 460 * 480	ACACATCTCCTCCATACTCAGGACTCCTCTCCGAGTTCCTGTCCCTTCTCTTAATTTAA	CCL4
	C-G-----T-----	CCL4L
* 500 * 520 * 540	TCTTTTTTATGTGCCGTGTTATTGTATTAGGTGTCATTTCATTATTTATATTAGTTTAA	CCL4
	---C-----T--A-----T-----	CCL4L
* 560 * 580 * 600	CCAAAGGATAAGTGTCCCTATGGGGATGGTCCACTGTCAGTCTTCTCTGCTGTTGCAA	CCL4
	-----C-----	CCL4L
* 620 * 640 * 660	ATACATGGATAACACATTTGATTCTGTGTGTTTTCAATAAAAATTTAAAATAAAATGC	CCL4
	-----CG--A--CA-----	CCL4L
	AGACAGTTA	
	-----	

**Figura 3:** Aliniament de la seqüència de cDNA dels locus CCL4 i CCL4L començant des del lloc d'iniciació de la transcripció. Les diferències de seqüència es mostren com a nucleòtids no conservats. Per altra banda les diferències codificants estan ressaltades amb negre (així com el codó d'iniciació de la traducció) i els corresponents canvis d'aminoàcid (CCL4 → CCL4L) es detallen a sota. Les dues possibles senyals de poliadenilació estan subratllades.



**Figura 4:** Organització genòmica dels gens de CCL3 i CCL4 i els seus duplicats

## 1.5. VARIACIÓ GENÒMICA

Malgrat que aquest treball s'enmarca, de forma general, dins el camp de la immunologia, conté nombroses incursions en el camp de la genètica i, concretament, en l'àrea de la variació genòmica. Per tant, creiem indicat introduir aquí alguns dels conceptes clàssics sobre aquest tema i actualitzar-ne d'altres que últimament han evolucionat de forma notable. En tot cas sempre es tracta d'una brevíssima pinzellada de termes genètics.

### 1.5.1. Glossari general

A continuació es mostra un petit glossari amb petits apunts de les formes de variació genòmica més freqüents que contribueixen a les diferències fenotípiques entre individus:

#### 1.5.1.1. SNPs (de l'anglès *Single Nucleotide Polymorphism*)

Com diu el seu nom, és una variació en la seqüència de DNA que afecta a un sol nucleòtid. En una població, cada SNP presenta una determinada freqüència de l'al·lel menor (**MAF**, de l'anglès *Minor Allele Frequency*), i és important ressaltar que, a l'existir fortes diferències entre les poblacions humanes, un al·lel que és comú en una determinada àrea geogràfica, pot ser rar en d'altres. L'anàlisi de milers d'SNPs del genoma humà apunta a la prevalença d'una estructura general en blocs, on certes regions semblen ser punts calents de recombinació. Aquests punts calents delimiten blocs haplotípics on la recombinació és rara o pràcticament absent i on s'observa un elevat desequilibri de lligament (LD) i una baixa diversitat haplotípica.

#### 1.5.1.2. Repeticions nucleotídiques

Dues són les denominacions dels principals tipus de repeticions nucleotídiques, **Microsatèl·lits** i **Minisatèl·lits**.

Sovint també descrits com a STRs (de l'anglès *Simple/Short Tandem Repeats*), els **microsatèl·lits** apareixen com a conseqüència de repeticions curtes de nucleòtids (típicament unitats d'un a quatre nucleòtids repetits de 10 a 20 vegades) que trobem distribuïdes al llarg de tot el genoma. Entre les repeticions de mononucleòtids, les que impliquen A o T són les més freqüents (sent molt menys usuals les de G o C). En el cas de les repeticions variables de dinucleòtids, les més freqüents són les CA (TG a la cadena complementària). També trobem sovint repeticions CT/AG i són força rares la variabilitat de les CG/GC. Les repeticions de trinucleòtids i tetranucleòtids són comparativament força menys freqüents, però sovint són altament polimòrfiques.

D'altra banda emprem el terme **minisatèl·lits** quan les repeticions afecten de 10 a 100 nucleòtids (malgrat que el rang més freqüent és de 10 a 25 nucleòtids); generalment, els minisatèl·lits són rics en G/C.

Aquestes repeticions (micro i minisatèl·lits) solen tenir variabilitat en el nombre de còpies definint diferents al·lels en una població. Aquesta variabilitat al·lèlica en les repeticions es coneix amb el nom de **VNTRs** (de l'anglès *Variable Number of Tandem Repeats*); tot i que la variabilitat pot afectar tant a microsatèl·lits com a minisatèl·lits, el terme VNTR es sol reservar per als minisatèl·lits, com a contraposició del terme STR, habitualment emprat en sinonímia amb microsatèl·lits.

### 1.5.1.3. Altres elements de variabilitat genòmica: Insercions/Delecions, Inversions i Translocacions

Les **insercions** (i delecions) poden ser de dos tipus: nucleotídiques discretes o cromosòmiques. Les primeres (discretes) són adicions (o pèrdues) d'un o més nucleòtids en una determinada seqüència de DNA. Sovint succeeixen en regions amb microsatèl·lits degut a errors de la DNA polimerasa. Quan ens hi referim a nivell cromosòmic, el terme inserció es refereix a l'adició d'una seqüència llarga en un cromosoma (habitualment degut a un entrecreuament desigual durant la meiosi) i, de la mateixa manera, les **delecions** succeeixen quan una part d'un cromosoma o una seqüència de DNA es perd (les delecions a nivell cromosòmic també poden ser causades per entrecreuaments erronis durant la meiosi).

Les **inversions** són reordenaments cromosòmics en els quals un segment d'un cromosoma s'inverteix respecte a la resta de regions genòmiques que l'envolten. N'hi ha de dos tipus: Paracèntriques (quan ambdós punts de trencament succeeixen en un sol braç del cromosoma; no inclouen el centròmer) i pericèntriques (quan trobem un punt de trencament a cada braç; inclouen per tant el centròmer).

Una **translocació** cromosòmica és una anormalitat causada per un reordenament de segments entre cromosomes no homòlegs. Les translocacions poden ser equilibrades o no equilibrades (segons si en el procés es manté o no intacta la informació genètica).

### 1.5.2. Variants estructurals

Quan parlem de variacions estructurals en el genoma ens referim a aquelles variants citogenèticament visibles o, més freqüentment, submicroscòpiques que inclouen delecions, insercions, duplicacions i variants de nombre de còpies a gran escala [collectivament anomenades Variacions de Nombre de Còpies (CNV)], així com inversions i translocacions. Per consens, es consideren variants estructurals aquelles que involucren segments de DNA més grans d'una Kb. La variació estructural pot ser **quantitativa** (com les CNV que inclouen delecions, insercions i duplicacions) i/o **posicional** (translocacions) o **orientacional** (inversions). Concretament, ens referim a **CNV** quan parlem d'un segment de DNA de, com a mínim, 1 Kb, pel qual s'han observat diferències en el seu nombre de còpies en la comparació de dos o més genomes. Fins al moment, el concepte CNV no implica cap freqüència mínima o efecte fenotípic. Paral·lelament, s'anomena polimorfisme de nombre de còpies (**CNP**, de l'anglès *Copy Number Polymorphism*) a una CNV que es trobi en més d'un 1% de la població [110].

Fins fa poc, es consideraven els SNPs com la principal forma de variació genòmica i se'ls feia responsables de la major part de la variació fenotípica normal. Però recentment diversos grups han descrit l'àmplia i generalitzada presència de variacions de nombre còpia en individus normals [111, 112]. Arrel d'aquesta nova perspectiva, s'ha revisat el coneixement de les variacions de nombre de còpies i la seva significació en la variació fenotípica humana, incloent el seu paper en la resistència/susceptibilitat a diverses malalties [113-120]. D'entre els treballs que s'han publicat en aquest sentit, un dels més recents analitza el possible paper del nombre de còpies de *CCL3L* en l'artritis reumatoide [108]. Els resultats mostren que un nombre de còpies superior a 2 representa un factor de risc moderat (OR=1,34 p=0,009) per a l'artritis reumatoide en una cohort de pacients de Nova Zelanda (n=834) però no en una altra cohort de pacients del Regne Unit (n=302, p=0,643). La diferència de resultats entre les dues cohorts podria representar diferències genètiques o ambientals entre les dues poblacions. Per altra banda, el menor nombre de mostres de la cohort del Regne Unit disminueix el poder per a detectar l'associació. Aquest tipus d'inconsistència també s'ha vist en el cas de la relació de *CCL3L-CCL4L* amb la infecció per HIV. Mentre que inicialment es va descriure una marcada influència del nombre de còpies d'aquesta regió cromosòmica en la malaltia [109] (en els HIV+ hi ha menys còpies), un treball posterior mostra la seva manca d'impacte en aquesta infecció [121].

En qualsevol cas, sembla evident que l'era dels estudis que involucren CNVs en la resistència/susceptibilitat a diverses malalties està tan sols en els inicis i que cal aprofundir sobre aquest tema per tenir una visió clara del paper d'aquest tipus de variació estructural en patologia.





## ***2. HIPÒTESI***



## 2. HIPÒTESI

La complexitat que existeix en la xarxa de relacions que formen les quimiocines es deu, en bona part, al constant solapament funcional entre els seus membres fruit dels diferents sistemes de generació de variants altament homòlogues.

En base a aquest concepte descrivim la hipòtesi d'aquest treball per la qual postulem que els mecanismes evolutius de generació de diversitat (processos de duplicació i fusió gènica, tall-i-unió (*splicing*) alternatiu i proteolisi post-traducciona) que han contribuït a la formació de l'actual superfamília de quimiocines humanes, es reflecteixen en l'organització genòmica dels seus membres i, per tant, l'anàlisi d'aquesta organització genòmica i d'aquests mecanismes ens ha de permetre establir millor les diferents relacions funcionals entre ells i definir patrons de susceptibilitat a malaltia. Alhora, com a model experimental concret, aquest treball es centra en les quimiocines CCL4 i CCL4L en les que postulem que s'hi plasmen els múltiples mecanismes de generació de variabilitat de les quimiocines.



***3. OBJECTIUS***



### 3. OBJECTIUS

Basats en la variabilitat de *CCL4* i *CCL4L* i en la revisió global de la variabilitat a la superfamília de les quimiocines, els objectius concrets plantejats en aquest treball són els següents:

A) Basats en la variabilitat de *CCL4* i *CCL4L*:

1. Caracteritzar les variants derivades dels locus *CCL4* i *CCL4L* mitjançant fenòmens de tall-i-unió alternatiu i polimorfismes. Analitzar quantitativament l'expressió d'aquestes variants a nivell d'RNA missatger. Valorar la incidència de les variants en patologia.
2. Definir quines són les confusions més habituals i rellevants en les seqüències de *CCL4* i *CCL4L* presents en els catàlegs d'algunes empreses subministradores: clarificar l'origen de les seqüències i remarcar les diferències de cada variant per tal d'evitar l'ús de productes erronis.
3. Determinar la distribució poblacional mundial del nombre de còpies de *CCL4L* i dels seus SNPs més rellevants, utilitzant un protocol que ens permeti discriminar el nombre exacte de còpies de cada al·lel en individus heterozigots amb més de 2 còpies de *CCL4L*. Establir si existeix correlació entre el nombre de còpies de *CCL4L* i les freqüències dels SNPs.

B) Basats en la revisió global de la variabilitat a la superfamília de les quimiocines:

4. Realitzar una mapa exhaustiu de l'organització genòmica dels membres de la superfamília de les quimiocines i establir connexions entre aquesta organització genòmica i les seves funcions.
5. Ampliar el coneixement de la complexitat de la superfamília de les quimiocines mitjançant la recopilació dels polimorfismes més rellevants dels seus membres i de les formes generades per tall-i-unió alternatiu. Establir la relació d'aquests fenòmens amb patologia.





## ***4. RESULTATS***



## 4.1. CAPÍTOL I

### OBJECTIU:

Caracteritzar les variants derivades dels locus *CCL4* i *CCL4L* mitjançant fenòmens de tall-i-unió alternatiu i polimorfismes. Analitzar quantitativament l'expressió d'aquestes variants a nivell d'RNA missatger. Valorar l'incidència de les variants en patologia.

### ARTICLE:

#### **Multiple Products Derived from Two CCL4 Loci: High Incidence of a New Polymorphism in HIV+ Patients.**

Roger Colobran, Patricia Adreani, Yaqoub Ashhab, Anuska Llano, José A. Esté, Orlando Dominguez, Ricardo Pujol-Borrell, and Manel Juan.

*The Journal of Immunology*, **2005**, 174: 5655–5664.



#### 4.1.1. CONSIDERACIONS PRÈVIES SOBRE L'ARTICLE

Quan es va iniciar l'estudi que ha donat lloc a aquest article i, en definitiva, a aquesta tesi doctoral, el nostre laboratori ja portava uns anys interessat en el paper de les quimiocines en la resposta immunitària, especialment en una malaltia autoimmunitària de la tiroide com és la malaltia de Graves-Basedow. En un article publicat poc abans de l'inici de la tesi, es demostrava la sobreexpressió de les quimiocines inflamatòries CCL3 i CCL4 en glàndules de pacients amb la malaltia de Graves-Basedow [122] i, durant l'anàlisi dels trànscripats procedents de les diverses quimiocines estudiades, s'havia observat que, en el cas de CCL4, no només s'observaven trànscripats procedents dels dos locus (CCL4 i CCL4L), sinó que es va detectar la presència d'algunes seqüències corresponents a aquesta quimiocina però amb variacions importants. El fet de comprovar que el segon locus de CCL4 s'expressava va despertar el nostre interès ja que cal tenir en compte que, malgrat que l'existència d'un segon locus de CCL4 s'havia ja descrit en el moment del descobriment d'aquesta quimiocina (any 1990) [103], el seu paper en la resposta immune no havia estat estudiat. Això és degut a que, al principi, no es va poder demostrar la seva expressió i, fins i tot, durant un temps es va especular que pogués ser un pseudogen. Per tant, fins aquell moment (any 1999) els treballs publicats que incloïen l'estudi de CCL4, tenien només en compte el locus originalment descrit. De fet, aquesta dinàmica s'ha mantingut fins l'actualitat i pocs estudis tenen en compte el locus CCL4L quan s'interessen per CCL4.

#### 4.1.2. RESUM DE DADES DEL PRIMER CAPÍTOL:

Amb aquest panorama, l'objectiu general que inicialment es va definir per a aquest treball de tesi va ser l'aprofundir en el coneixement del paper de CCL4L en la resposta immune, caracteritzant primer el seu perfil d'expressió transcripcional (en comparació al de CCL4) i plantejant posteriorment la producció de la proteïna per a realitzar estudis funcionals i veure possibles diferències entre CCL4 i CCL4L. Aquest plantejament inicial va alterar-se substancialment per l'aparició de noves variants de CCL4 i CCL4L. Així en la primera part d'aquest capítol, es presenta la descripció detallada d'aquestes variants i els mecanismes que les generen. El primer resultat a valorar és que tan CCL4 com CCL4L expressen, juntament amb el trànscrip complet, trànscripats generats per un tall-i-unió alternatiu del mRNA als quals els manca l'exó 2. Aquest descobriment succeí a l'inici de l'estudi, on un dels primers requisits per estudiar de forma conjunta però independent CCL4 i CCL4L era poder amplificar-les específicament. Això, degut a l'elevada homologia nucleotídica dels dos locus (>95%, veure figura 3), va comportar certes dificultats a nivell de disseny d'oligonucleòtids. Durant l'optimització d'aquest procés van aparèixer uns trànscripats no esperats que, després de seqüenciar-los, van resultar ser aquestes

formes generades per un fenomen de tall-i-unió alternatiu. Vam anomenar a aquestes variants de CCL4 i CCL4L mancades de l'exó 2 com a CCL4 $\Delta$ 2 i CCL4L $\Delta$ 2 (on s'utilitza la lletra  $\Delta$  per indicar deleció/pèrdua i el nombre 2 com a indicador de que es tracta de l'exó 2).

Posteriorment, durant l'amplificació del gen *CCL4L* en mostres procedents de diferents individus, sorgiren alguns casos on el patró de trànscripats amplificats no corresponia en absolut amb el teòricament previst. La seqüenciació del DNA genòmic d'aquestes mostres va manifestar un canvi nucleotídic en el lloc acceptor de tall de l'intró 2 del gen *CCL4L*, concretament, en la primera posició del dinucleòtid AG, que representa la seqüència consens de tall. Aquest nou polimorfisme té unes conseqüències força dràstiques ja que fa desaparèixer la seqüència de tall original de l'intró 2 i possibilita l'ús de diverses noves seqüències de tall alternatives que generen un seguit de nous trànscripats. Per diferenciar les dues variants al·lèliques que genera aquest nou polimorfisme vam anomenar com a *CCL4L1* el gen originalment descrit i com a *CCL4L2* el gen amb el canvi nucleotídic (a l'article es descriuen perfectament totes les noves variants que origina *CCL4L2*). Cal remarcar que aquest polimorfisme de nucleòtid únic (o SNP) que es descriu a l'article com a +590 (A/G), correspon actualment al símbol rs4796195.

L'aparició de totes aquestes variants (tant les originades pel fenomen de tall-i-unió alternatiu com les derivades del nou polimorfisme) va augmentar notablement la complexitat de la parella CCL4-CCL4L. El nou panorama es descriu en la figura 4 de l'article, on podem veure des de l'organització genètica de *CCL4-CCL4L*, fins a la generació de les proteïnes potencials de totes les variants.

La segona part de l'article, analitza l'expressió de CCL4 i CCL4L, tenint en compte les variants descrites anteriorment. Destacar que la contribució general del locus *CCL4L* a l'expressió total de CCL4-CCL4L és de prop del 50%, indicant la importància que pot tenir l'anàlisi d'ambdues quimiocines, i no només CCL4, en moltes situacions inflamatòries.

La tercera i última part de l'article és un estudi cas-control per veure la possible incidència del polimorfisme en patologia. El resultat més destacable és l'evidència que l'al·lel *CCL4L2* està significativament augmentat en els pacients HIV<sup>+</sup> respecte els controls sans.

# Multiple Products Derived from Two CCL4 Loci: High Incidence of a New Polymorphism in HIV<sup>+</sup> Patients<sup>1</sup>

Roger Colobran,<sup>\*‡</sup> Patricia Adreani,<sup>\*</sup> Yaqoub Ashhab,<sup>\*</sup> Anuska Llano,<sup>†</sup> José A. Esté,<sup>†</sup> Orlando Dominguez,<sup>\*</sup> Ricardo Pujol-Borrell,<sup>\*‡</sup> and Manel Juan<sup>2\*‡</sup>

Human CCL4/macrophage inflammatory protein (MIP)-1 $\beta$  and CCL3/MIP-1 $\alpha$  are two highly related molecules that belong to a cluster of inflammatory CC chemokines located in chromosome 17. CCL4 and CCL3 were formed by duplication of a common ancestral gene, generating the *SCYA4* and *SCYA3* genes which, in turn, present a variable number of additional non-allelic copies (*SCYA4L* and *SCYA3L1*). In this study, we show that both CCL4 loci (*SCYA4* and *SCYA4L*) are expressed and alternatively generate spliced variants lacking the second exon. In addition, we found that the *SCYA4L* locus is polymorphic and displays a second allelic variant (hereinafter *SCYA4L2*) with a nucleotide change in the intron 2 acceptor splice site compared with the one described originally (hereinafter *SCYA4L1*). Therefore, the pattern of *SCYA4L2* transcripts is completely different from that of *SCYA4L1*, since *SCYA4L2* uses several new acceptor splice sites and generates nine new mRNAs. Furthermore, we analyzed the contribution of each locus (*SCYA4* and *SCYA4L1/L2*) to total CCL4 expression in human CD8 T cells by RT-amplified fragment length polymorphism and real-time PCR, and we found that L2 homozygous individuals (L2L2) only express half the levels of CCL4 compared with L1L1 individuals. The analysis of transcripts from the *SCYA4L* locus showed a lower level in L2 homozygous compared with L1 homozygous individuals (12% vs 52% of total CCL4 transcripts). A possible clinical relevance of these CCL4 allelic variants was suggested by the higher frequency of the L2 allele in a group of HIV<sup>+</sup> individuals ( $n = 175$ ) when compared with controls ( $n = 220$ , 28.6% vs 16.6% ( $p = 0.00016$ )). *The Journal of Immunology*, 2005, 174: 5655–5664.

Chemokines are a large superfamily of small (~8–15 kDa) structurally related molecules that regulate cell trafficking of various types of leukocytes to areas of injury or infection and play different roles in both inflammatory and homeostatic processes (1–3). According to the arrangement of a structural cysteine motif found near the NH<sub>2</sub> terminus of the mature protein, chemokines have been divided into four subfamilies: CXC, CC, CX<sub>3</sub>C, and C chemokines (4). These chemokines carry out their biological functions through seven-transmembrane domain G protein-coupled receptors, which are expressed on several populations of leukocytes (5–7). To date, 42 chemokines and 18 chemokine receptors have been identified in humans (8).

A remarkable feature of chemokines and chemokine receptors is their redundancy and binding promiscuity. There are many examples of a single chemokine binding to several receptors, as well as a single chemokine receptor transducing signals for several chemokines. Interestingly, genes encoding the inflammatory chemokines tend to appear in clusters (human CC subfamily in chromosome 17 and the CXC subfamily in chromosome 4). Moreover, an

important characteristic of cluster chemokines is that they share many ligands with few receptors (4). Individual members of the chemokine superfamily may contribute to this complex relationship network by two mechanisms of additional variability: 1) pretranslational modifications (such as alternative splicing) (9, 10) and 2) post-translational modifications (such as NH<sub>2</sub>-terminal truncations by the dipeptidyl-peptidase (DPP)<sup>3</sup> CD26/DPP IV) (11–14).

CCL3/macrophage inflammatory protein (MIP)-1 $\alpha$  and CCL4/MIP-1 $\beta$  are two highly related chemokines that belong to a cluster of inflammatory CC chemokines located in chromosome 17 (q11-q21), which are secreted by specific cells after being triggered by Ags or mitogenic signals and attract additional cells involved in immune responses (15). These two chemokines, together with CCL5/RANTES, are the major HIV-suppressive factors produced by CD8<sup>+</sup> T cells (16) by binding to CCR5, the coreceptor necessary for the entry of HIV-R5 strains into CD4<sup>+</sup> cells (17, 18).

CCL3 and CCL4 were formed by duplication of a common ancestral gene (15) that originated *SCYA3/LD78 $\alpha$*  and *SCYA4/ACT-2* genes which in turn have a second non-allelic copy (*SCYA3L1/LD78 $\beta$*  and *SCYA4L/LAG-1*) present in variable numbers in the human genome (19). In the case of CCL3, there is a third gene (*SCYA3L2/LD78 $\gamma$* ) that is a 5'-truncated pseudogene (20). Therefore, human CCL3 and CCL4 are encoded by two highly related non-allelic isoforms that have been duplicated and mutated to produce two different but highly homologous proteins (>90% between CCL3 (from *SCYA3*) and CCL3L1 (from *SCYA3L1*) proteins, and >95% between CCL4 (from *SCYA4*) and CCL4L (from *SCYA4L*) proteins) (21–23). In CCL3, functional differences have been reported between the proteins of two loci. CCL3L1/LD78 $\beta$  has been characterized as a more potent CCR5 agonist and a greater inhibitor of HIV-1 than CCL3/LD78 $\alpha$  (24–27). Moreover,

\*Laboratory of Immunobiology for Research and Application to Diagnosis, Centre for Transfusion and Tissue Bank, Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol, and †Retrovirology Laboratory, IrsiCaixa Foundation, Hospital Universitari Germans Trias i Pujol, ‡Department of Cell Biology, Physiology and Immunology, Universitat Autònoma de Barcelona, Barcelona, Spain

Received for publication October 13, 2004. Accepted for publication February 18, 2005.

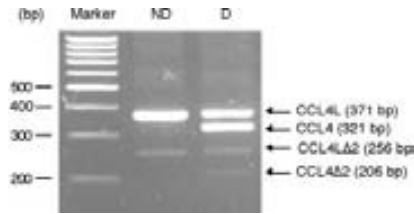
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> This work was supported by grants from the Fondo de Investigaciones Sanitarias (project FIS 99/1063 and PI 02/0104), the Ministerio de Educación y Ciencia (project BFI 2003–00405), and the Instituto de Salud Carlos III (RC03/03).

<sup>2</sup> Address correspondence and reprint requests to Dr. Manel Juan, Laboratory of Immunobiology for Research and Application to Diagnosis, Immunology Unit, Hospital Universitari Germans Trias i Pujol, Ctra. de Canyet s/n, 08916 Barcelona, Spain. E-mail address: mjuan@ns.hugtip.scs.es

<sup>3</sup> Abbreviations used in this paper: DPP, dipeptidyl-peptidase; MIP, macrophage inflammatory protein; GAG, glycosaminoglycan; SDF, stromal cell-derived factor.





**FIGURE 1.** Two canonical products of CCL4 and their variants lacking exon 2. cDNA from stimulated CD8 T cells was amplified using primers in exons 1 and 3. The PCR product was not digested (ND) or digested (D) with *MspI*. The highest m.w. amplicon in lane ND contains CCL4L and CCL4 products resolved after *MspI* restriction analysis (lane D). The lowest m.w. amplicon in lane ND contains CCL4LΔ2 and CCL4Δ2 products resolved after *MspI* digestion (lane D).

the potent antiviral activity of CCL3L1/LD78β is increased due to the cleavage by CD26/DPP IV, a dipeptidyl-peptidase that cuts dipeptides from the NH<sub>2</sub> terminus of regulatory peptides with a proline or alanine residue in the penultimate position (14, 28). Recently, CCL4 has also been described as a target for CD26/DPP IV, being mainly secreted by stimulated PBLs as a truncated form that lacks the first two amino acids. This posttranslational modification affects receptor specificity but not anti-HIV activity (29, 30).

Our studies point out that, in fact, the word “CCL4” groups a broad quantity of different transcripts coded by the *SCYA4* and *SCYA4L* loci. This high variability is caused by an alternative splicing mechanism and by the presence of a polymorphism in the *SCYA4L* locus that changes the intron 2 acceptor splice site and creates different mRNA expression patterns. We have assessed the possibility that individuals carrying different genotypes and with different capabilities to exert this diversity of CCL4 variants may respond differently in situations of chronic immunostimulation, such as viral infection or autoimmunity. Therefore, the possible clinical relevance of these CCL4 allelic variants was suggested by

the higher frequency of the *SCYA4L2* allele in a group of HIV<sup>+</sup> individuals when compared with controls.

**Materials and Methods**

*Isolation and stimulation of human CD8 T cells*

PBMCs were isolated from total blood by centrifugation over Lymphoprep (Axis-Shield). CD8 T cells were positively selected using magnetic beads coated with mAbs against CD8 (MACS MicroBeads; Miltenyi Biotec) according to the manufacturer’s instructions. Purity of CD8 T cells was routinely >98%, as determined by flow cytometry.

CCL4 production by purified CD8 T cells was measured at baseline and after PHA stimulation (31). Cells were cultured in RPMI 1640 complete medium (Invitrogen Life Technologies) containing 10% FCS, streptomycin, and penicillin at a cell concentration of 1 × 10<sup>6</sup> cells/ml in 24-well plates (Corning Costar). To induce CCL4 production, we added 1 μg/ml PHA and 10 ng/ml recombinant human IL-2 (R&D Systems). Cells were harvested and total RNA was extracted after 0, 6, 24, and 48 h of PHA stimulation. Two replicate experiments were performed for each condition.

*Human CCL4 amplification*

CCL4 was amplified by RT-PCR. RNA was extracted using the RNAqueous-96 kit (Ambion) and was then retrotranscribed with oligo(dT)<sub>15</sub> and SuperScript-II (Amersham Biosciences). PCR was performed in 10 μl of total reaction volume. The primers used to amplify all CCL4 variants (from both loci) were 82A/Bex1 5’-GAAGCTCTGCGTGACTGTC-3’ and 433A/Bex3 5’-CGGAGAGGAGTCTGAGTAT-3’, and the PCR cycling conditions were 95°C for 30 s, 62°C for 30 s and 72°C for 30 s. The primers used to amplify *SCYA4L* variants were 124Bex1 5’-TGCCTTCTGCTCTAGCA-3’ and 615Bex3 5’-TGAAAACACATGGAATTAACG-3’, and the PCR cycling conditions were 95°C for 30 s, 59°C for 30 s and 72°C for 30 s. PCR products were visualized by ethidium bromide staining following electrophoresis in 2% Tris borate-EDTA (TBE) high-resolution agarose gels (Sigma-Aldrich).

*Restriction digests and densitometric analysis*

PCR products were digested with 10 units of *MspI* (Fermentas) in 15 μl of total reaction volume for 90 min at 37°C. The digested products were evaluated by visualization by ethidium bromide staining following electrophoresis in 2% TBE high-resolution agarose gels. PCR products and subsequent restriction fragments were quantified using the Quantity One software (Bio-Rad) according to the manufacturer’s instructions.



**FIGURE 2.** Alignment of *SCYA4* and *SCYA4L* cDNA with their amino acid sequences. Lanes *SCYA4* and *SCYA4L* are cDNAs, while CCL4/CCL4L are their canonical peptide products and CCL4Δ2/CCL4LΔ2 are their variants lacking exon 2. Conserved nucleotides are shown as dashes. Exons are separated with vertical bars. Absent amino acids in the variants lacking exon 2 are shown as asterisks (\*), and amino acid differences are bold. Amino acids of the signal peptide are in lowercase letters whereas amino acids of the mature protein are in uppercase letters. The initiation codon, the stop codons, and two possible polyadenylation signals are shaded with black. AU-rich sequences that reduce mRNA stability are shaded in gray. The *MspI* restriction target is underlined.

**A**

Gene name	Variant name	Accession number	mRNA	Mature protein	Sequence of 5' splice site	Position of 5' splice site	Score (0-100)
SCYA4	CCL4	AY766466	667 bp	69 aa	TCCTTGTTCTACAGA	Canonical	91.6
SCYA4	CCL4Δ2	AY766448	552 bp	29 aa	TCCTTGTTCTACAGA	Canonical	91.6
SCYA4L1	CCL4L	AY766447	667 bp	69 aa	TCCTTGTTCTACAGA	Canonical	91.6
SCYA4L1	CCL4LΔ2	AY766449	552 bp	29 aa	TCCTTGTTCTACAGA	Canonical	91.6
SCYA4L2	CCL4L2	AY766450	652 bp	64 aa	TTCCAAACCCAAAGA	Exon 3 (+15)	66.4
SCYA4L2	CCL4L2Δ2	AY766456	537 bp	24 aa	TTCCAAACCCAAAGA	Exon 3 (+15)	66.4
SCYA4L2	CCL4L2b2	AY766452	1258 bp	41 aa	-----	No splicing	---
SCYA4L2	CCL4L2b1	AY766451	856 bp	41 aa	CCTGGGGCCACAGC	Intron 2 (-189)	74.9
SCYA4L2	CCL4L2bΔ2	AY766457	741 bp	45 aa	CCTGGGGCCACAGC	Intron 2 (-189)	74.9
SCYA4L2	CCL4L2f	AY766458	846 bp	80 aa	CAGCTAAATCCCAAGT	Intron 2 (-179)	70.9
SCYA4L2	CCL4L2c	AY766453	836 bp	80 aa	CCAGTGAAGTGGAAAGT	Intron 2 (-169)	53.6
SCYA4L2	CCL4L2d	AY766454	815 bp	73 aa	GAGTCTGCTTCCCAAGT	Intron 2 (-148)	80.3
SCYA4L2	CCL4L2e	AY766455	785 bp	63 aa	ATCCCATCCACCAGA	Intron 2 (-118)	81.2

**B**

**FIGURE 3.** Multiple transcripts derived from a polymorphism in *SCYA4L* locus. *A*, General description of all detected *CCL4* mRNAs. *SCYA4L1*-derived mRNAs are shaded in black and *SCYA4L2*-derived mRNAs are shaded in gray. The score of acceptor splicing sites is based on Shapiro and Senepathy (35). *B*, Scheme of splicing in *SCYA4* and *SCYA4L1/SCYA4L2* genes. In *SCYA4L2*, the original acceptor site is mutated (AG > GG) and the spliceosome is unable to recognize it. Instead, the spliceosome can select another alternative acceptor splice site around the original site. Accession numbers correspond to those provided by GenBank after submission. Gene accession numbers: *SCYA4*, AY766459; *SCYA4L1*, AY766460; *SCYA4L2*, AY766461.

### Population study of *SCYA4L* locus polymorphism

To evaluate the population frequencies of *SCYA4L* polymorphism, we analyzed 220 healthy donors as well as 175 HIV-seropositive, 80 hepatitis C virus-seropositive, 30 insulin-dependent diabetes mellitus, and 30 autoimmune thyroid disease patients. All individuals are ethnically grouped as Caucasoid Spanish population. The genomic DNA was obtained from total blood of each individual, and by using the primers I2B 5'-GCAGAGGAA GATGCCTACCAC-3' and 503Bex3 5'-AAATAATGGAAATGACAC CTAATAC-3', we amplified the junction between intron 2 and exon 3 of the *SCYA4L* locus under the following PCR cycling conditions: 95°C for 30 s, 59°C for 30 s, and 72°C for 30 s. The PCR product was purified using the GFX PCR DNA and Gel Band purification kit (Amersham Biosciences). Finally, DNA sequencing was performed with both primers on an Applied Biosystems (ABI) Prism 3100 genetic analyzer.

### Software tools

To assess the influence of the change in the acceptor splice site sequence in the *SCYA4L2* allelic variant, we have used two software applications: 1) GENSCAN (<http://genes.mit.edu/GENSCAN.html>) (32) and 2) HMMgene (v. 1.1) (<http://www.cbs.dtu.dk/services/HMMgene/>) (33).

We used NNSPLICE (v. 0.9) ([http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html)) (34) to predict the potential acceptor splice sites in the *SCYA4L2* allelic variant.

The algorithm to calculate the scores of alternative splice sites found in the *SCYA4L2* allelic variant is based on Shapiro and Senepathy (35) (available at <http://www.genet.sickkids.on.ca/~ali/splicesitescore.html>).

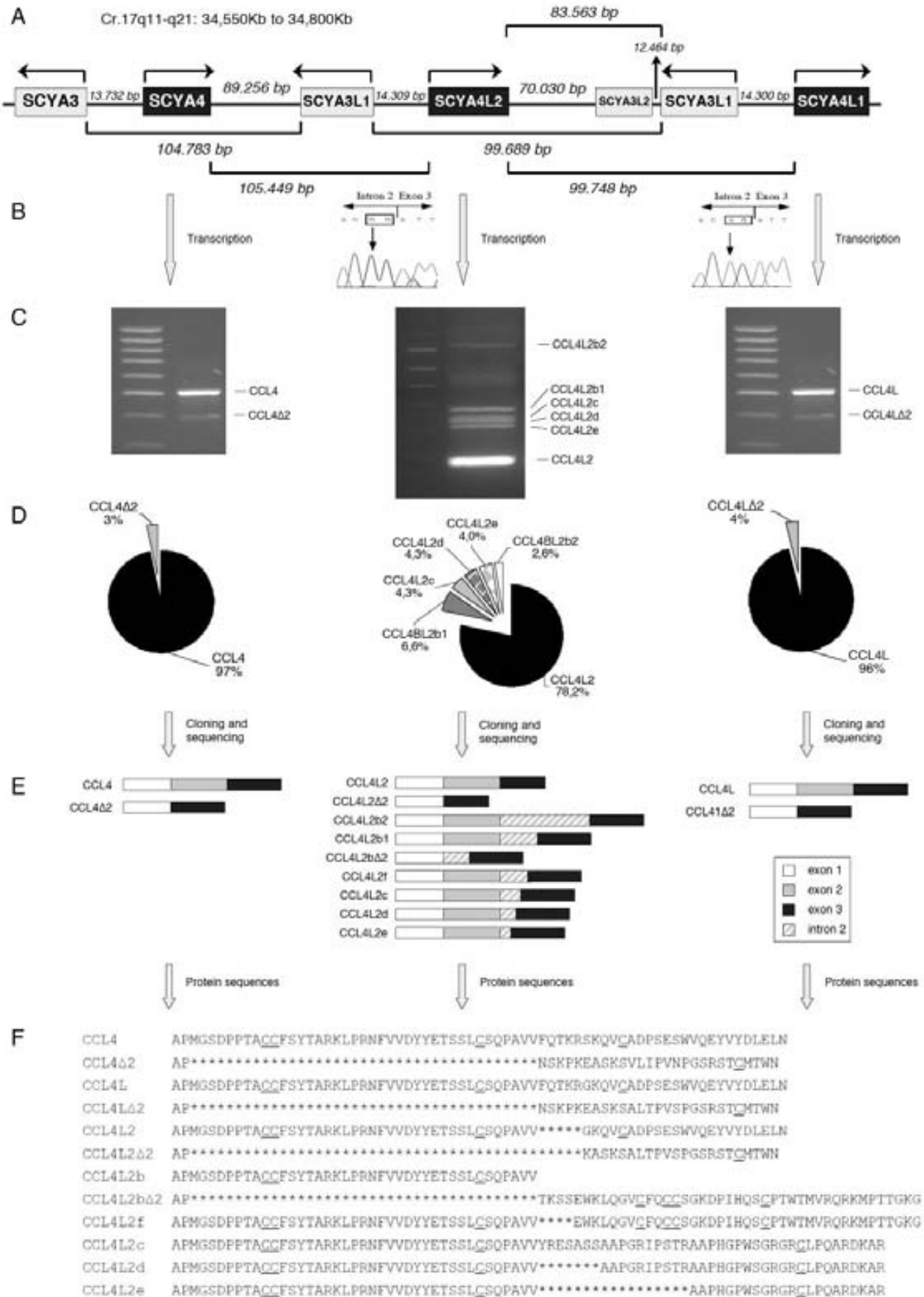
A genomic sequence from GenBank was used to establish the distribution and distances among *CCL3* and *CCL4* gene loci and their duplicates: NT\_010799 includes a 9,412,828-bp sequence referred to *Homo sapiens*

chromosome 17. This DNA sequence is part of the second release of the completed human reference genome. It was assembled from individual clone sequences by the Human Genome Sequencing Consortium in consultation with National Center for Biotechnology Information (NCBI) staff.

### Real-time PCR and expression studies

To evaluate the total expression of *CCL4*, we used real-time PCR with primers amplifying all potential variants (from *SCYA4* and *SCYA4L* loci). Standards for *CCL4* and *GAPDH* were obtained by conventional PCR from PHA-stimulated CD8 T cells cDNA. Amplification products were quantified in serial dilutions from  $10^8$  to  $10^1$  molecules. Real-time PCR from cDNA was performed in a LightCycler (Roche Diagnostics) using the master mix containing 4 mM  $MgCl_2$ , 0.5  $\mu$ M primers (82A/Bex1 and 433A/Bex3), and 1  $\mu$ l of LightCycler Fast Start DNA Master SYBR Green I (Roche Diagnostics). The amount of cDNA was calculated using the second derivate method after confirming the specificity of the amplification with the melting curve profiles. Two replicates of each sample were performed for *CCL4* and *GAPDH*, and a maximum SD of 15% between replicates was accepted. The relative abundance of *CCL4* in each sample was calculated by normalizing the mean levels of *CCL4* mRNA copies ( $mean_{CCL4-sample}$ ) with the corresponding mean value for *GAPDH* ( $mean_{GAPDH-sample}$ ) using the formula  $index_{sample} = (mean_{CCL4-sample}) / (mean_{GAPDH-sample})$ .

To analyze the contribution of *SCYA4* and *SCYA4L* loci-derived variants to the total *CCL4* mRNA, real-time PCR products were recovered and half of the total amount was digested with *MspI* as described. The digested and non-digested products were visualized in a high-resolution agarose gel electrophoresis and quantified by densitometry as described.



### Statistics

Data sets were analyzed using SPSS v. 11.0.1 for Macintosh (SPSS). When necessary, the results were expressed as the mean value  $\pm$  SD. An independent-samples Student's *t* test and a  $\chi^2$  test were applied to the data sets to determine statistically significant differences between groups. Differences were considered significant when *p* values were  $<0.05$ .

## Results

### The two *CCL4* loci generate four mRNAs

CCL4 was amplified from cDNA of human activated CD8 T cells using primers placed in exons 1 and 3. Besides the expected canonical product of the *SCYA4* and *SCYA4L* loci (371 bp), an unexpected low m.w. amplicon was observed (256 bp; Fig. 1, *ND lane*). The m.w. difference was 115 bp, which was in exact accordance with the length of the second exon of the loci. Cloning and sequencing these two amplicons showed that the 371-bp product included two different sequences corresponding to the CCL4 and CCL4L variants, whereas the unexpected low m.w. band included alternatively spliced variants of CCL4 and CCL4L, thus confirming the lack of exon 2 in 256-bp product. Therefore, the described CCL4 chemokine is codified by two loci that allow at least four different spliced forms: 1) two long, mature mRNAs (full length of 667 bp), with identical size for both loci (codifying for CCL4 and CCL4L proteins) but with 26 nucleotide differences (19 transitions and 7 transversions) and three amino acid changes, two in the signal peptide (V12M and L20P) and the third one (G70S) in the mature protein (Fig. 2); 2) two short spliced mRNAs (full length of 552 bp) also derived from both loci, but lacking the second exon. The predicted amino acid sequences of both forms present a frame shift in the exon 3 reading frame caused by the new junction between exon 1 and 3, and a stop codon appears close to the original one (Fig. 2). Consequently, these mature proteins are shorter (29 amino acids) and only maintain the two initial residues of the complete CCL4 and CCL4L proteins (69 amino acids). Interestingly, due to the frame shift, some of the conservative base substitutions between the two loci become non-conservative in these forms: there are three amino acid differences in the predicted mature protein (A13V, T15I, and S18N). Following commonly accepted nomenclature usage, they will be named CCL4 $\Delta$ 2 and CCL4L $\Delta$ 2, respectively.

### Multiple mRNAs derived from a new allelic variant (*SCYA4L2*) of *SCYA4L* locus

When we specifically amplified the *SCYA4L*-derived variants (CCL4L and CCL4L $\Delta$ 2) from different individuals, some samples showed an unexpected new pattern of unidentified amplicons. Cloning and sequencing these variants revealed nine new mRNAs formed by the use of different cryptic acceptor splice sites located in the second intron or in the third exon (Fig. 3A, shaded in gray). By sequencing the corresponding genomic DNA of these individuals, we discovered a +590A > G critical change in the original acceptor splice site of the second intron of the *SCYA4L* locus.

Therefore, the *SCYA4L* locus is a polymorphic gene with two allelic variants: the original *SCYA4L* locus (hereinafter *SCYA4L1*) and a second highly related polymorphic variant (hereinafter

*SCYA4L2*); despite the lack of differences between the three exons and the first intron of *SCYA4L1* and *SCYA4L2*, we have found four base substitutions in the second intron of *SCYA4L2*. Only the fourth of these changes (described above as +590A > G) seems to be critical for the final expression of *SCYA4L2*. Normally (36, 37), the donor splice site of the second intron in *SCYA4L1* shows GT right after the point where exon 2 finishes, whereas the acceptor site has AG just before the point where intron 2 sequence is cleaved (canonical pattern). In *SCYA4L2*, the sequence of the acceptor splice site in *SCYA4L1* experiences a critical change and becomes GG (Fig. 3B). According to the bioinformatic predictions, this polymorphism causes the inability of spliceosomes to recognize the mutated acceptor site (GG), and therefore two splicing options appear: 1) no removal of the second intron (hence generating a mature long mRNA with exon 1 + exon 2 + intron 2 + exon 3), or 2) use of alternative acceptor sites around the original one. Almost all mRNAs derived from *SCYA4L2* are generated following this second option (use of an alternative acceptor splice site), but we also observed a few mRNAs caused by lack of splicing in the second intron.

### *SCYA4* and *SCYA4L*: from 2 genes to 13 mRNAs

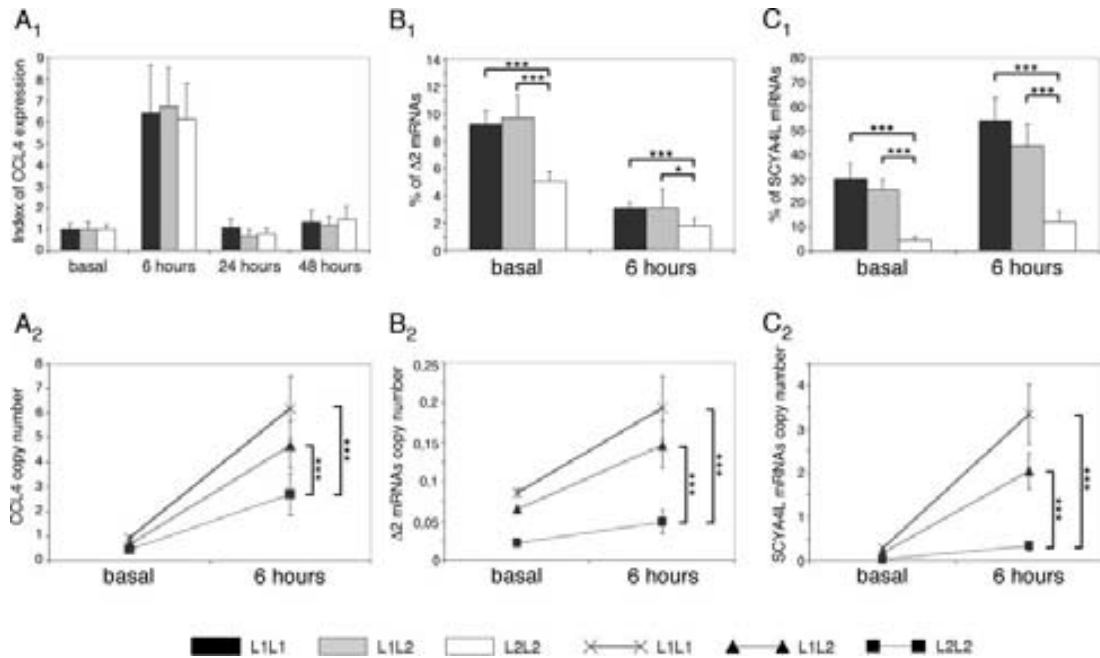
A search in the NCBI nucleotide database for CCL4 ([http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db = Nucleotide](http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Nucleotide)) provided access to a long stretch of the human genomic sequence in chromosome 17 (9,412,828 bp), assembled from individual clone sequences by the Human Genome Sequencing Consortium (NT\_010799). In this region we found not only the *SCYA3* and *SCYA4* loci head to head and separated by 13.7 kb, but also two copies of the alternative forms (*SCYA3L1* and *SCYA4L*) arranged in the same way and separated by a non-coding region of 14.3 kb (Fig. 4, physical map). Although mature mRNAs derived from *SCYA4* and *SCYA4L* have the same length (667 bp), differences in their respective introns resulted in different lengths of the primary transcripts (1795 bp in *SCYA4* vs 1816 bp in *SCYA4L*). A 12-bp deletion in intron 1 and an 11-bp deletion in intron 2 are the most relevant differences between the *SCYA4* and the *SCYA4L* loci.

Strikingly, the two copies of the *SCYA4L* locus present in this sequence are the two allelic variants described in this paper (*SCYA4L1* and *SCYA4L2*). Additional minor differences exist in the 5' and 3' non-coding regions of the two copies of *SCYA4L*, but we have not addressed their effect on transcription in this study.

*SCYA3* and *SCYA3L1*, as well as *SCYA4* and *SCYA4L2*, are separated by 105 kb. We also found the *SCYA3L2* pseudogene close to one of the copies of *SCYA3L1*. Other non-related loci (e.g., as a member of the *TRE17* oncogene family) are duplicated together with the *SCYA4L* and *SCYA3L1* alternative forms. All this suggests repeated duplications of a 120 kb stretch in this region of chromosome 17.

The most abundant mRNA derived from *SCYA4L2* (78.2% of total mRNA expression) corresponds to the new CCL4L2 variant of CCL4 generated by the use of an acceptor splice site located 15 nucleotides downstream of the original site (Fig. 4, densitometry of gel electrophoresis). The predicted CCL4L2 mature protein has 64 amino acids and lacks the initial five amino acids codified by

**FIGURE 4.** From genomic organization to mRNA products. *A*, Physical map from 34,550 kb to 34,800 kb, inside the 17q11-q21 region, based on the genomic sequence NT\_010799. *B*, Representative sequences of the junction between intron 2 and exon 3 from *SCYA4L1* and *SCYA4L2*, produced by amplification of genomic DNA from PBMCs by using specific primers for locus *SCYA4L*. Dinucleotides corresponding to original acceptor splice sites are squared, and the polymorphic position is indicated with an arrow. *C*, High-resolution agarose gel electrophoresis showing the transcription pattern of each CCL4 gene (cDNA from stimulated CD8 T cells was amplified using locus-specific primers). *D*, Percentages of expression of each form (SD is  $<5\%$  in each value). *E*, All mRNAs derived and cloned from *SCYA4* and *SCYA4L1/2* are represented with boxes. *F*, Alignment of derived mature proteins. Cysteines are underlined.



**FIGURE 5.** mRNA expression of CCL4 in CD8 T cells of L1L1, L1L2, and L2L2 individuals. *A*<sub>1</sub>, Time course of CCL4 mRNA expression. Expression levels are relative to the housekeeping gene (GAPDH), represented as an index (y-axis), and referred to the basal values (index = 1). *A*<sub>2</sub>, Copy number of CCL4 (relative to the GAPDH) in stimulated (6 h) and non-stimulated CD8 T cells. *B*<sub>1</sub>, Percentage of mRNAs lacking exon 2 compared with the total CCL4 mRNAs. *B*<sub>2</sub>, Relative copy number of mRNAs lacking exon 2. *C*<sub>1</sub>, Percentage of *SCYA4L* locus mRNAs relative to total CCL4 mRNAs. *C*<sub>2</sub>, Relative copy number of *SCYA4L*-derived mRNAs. Each point was performed in duplicate from three samples of different individuals, and the results were represented as the mean  $\pm$  SD. Differences between the basal and stimulated (6 h) situations were always statistically significant in all genotypes ( $p < 0.01$ ). \*,  $p < 0.05$ ; \*\*\*,  $p < 0.005$ .

the third exon (FQTKR), but the rest of the sequence remains unchanged (Fig. 4, protein sequence alignment).

We have named CCL4L2b, a 41-aa predicted form of CCL4 that is truncated by a premature stop codon and is codified by two mRNAs, CCL4L2b1 and CCL4L2b2, accounting for 6.6 and 2.6% of total message, respectively. The other CCL4 mRNAs detected (CCL4L2c, CCL4L2d, CCL4L2e, each accounting for around 4%) code for three new and highly related CCL4 variants that maintain the frame shift produced in the junction between exon 2 and the new exon 3 of each variant and use the same stop codon. Despite having a non-related amino acid sequence from position 42 to the end compared with the original CCL4L protein, these three proteins show a cysteine at -10 from the C terminus that may maintain the basic chemokine conformation.

Finally, we have cloned three very low frequency mRNAs (not visible by gel electrophoresis): 1) CCL4L2 $\Delta$ 2, which is the CCL4L2 variant lacking exon 2; 2) CCL4L2b $\Delta$ 2, which is the CCL4L2b variant lacking exon 2; and 3) CCL4L2f, which uses another different alternative acceptor splice site and has the same reading frame than CCL4L2b $\Delta$ 2. The CCL4L2f predicted protein (80 aa) presents seven cysteines in its sequence (with two CC motifs) (Fig. 4, protein sequences alignment).

In theory, other exon 2-skipped variants of all *SCYA4L2*-derived proteins (next to CCL4L2 $\Delta$ 2 and CCL4L2b $\Delta$ 2) may also exist, but we could not detect them due to their very low frequency.

#### Expression of CCL4 mRNA variants

The expression of CCL4 mRNA variants was assessed by real-time RT-PCR at 6, 24, and 48 h after PHA-stimulated and control CD8 T lymphocytes from healthy donors of each CCL4 genotype: L1L1, L1L2 and L2L2 (three individuals for each group). To evaluate the contribution of each locus to total mRNA expression, we took advantage of the presence of a *MspI* restriction site in the

products of *SCYA4* locus (38). Peak CCL4 expression was observed at 6 h with an average increase of 6- to 7-fold compared with controls in all genotypes (Fig. 5*A*<sub>1</sub>, data of 24 and 48 h are only shown in this figure and not in the following because mRNA quantities returned to levels similar to those obtained from non-stimulated samples). Interestingly, the ratio of CCL4 to GAPDH mRNA was higher in L1L1 than in L2L2 individuals (6.2 vs 2.7; Fig. 5*A*<sub>2</sub>). Values obtained from heterozygous were intermediate (see graphics in Fig. 5).

The ratio of short mRNAs lacking the second exon to total CCL4 mRNA was 9.2% in L1L1 individuals (which express the CCL4 $\Delta$ 2 and CCL4L $\Delta$ 2 variants), but only 5% in L2L2 individuals (where only CCL4 $\Delta$ 2 may be detected) (Fig. 5*B*<sub>1</sub>); these ratios decrease to 3.1 and 1.8%, respectively, after 6 h of stimulation with PHA, despite the increase in absolute values (Fig. 5*B*<sub>2</sub>).

Finally, we analyzed the relative contribution of the *SCYA4L* locus to total CCL4 expression. In basal conditions, *SCYA4L*-derived mRNAs accounted for 29.9% of total CCL4 expression, in L1L1 individuals, but only for 4.6%, in L2L2 individuals (Fig. 5*C*<sub>1,2</sub>). After stimulation for 6 h, the percentage of *SCYA4L* mRNAs was 54% in L1L1 individuals, and 12.1% in L2L2 subjects. Hence, activation seems to shift transcription to the *SCYA4L* locus in all genotypes. In both conditions, the contribution of *SCYA4L2* to the total levels of CCL4 mRNA is very low compared with *SCYA4L1*.

#### Population incidence of *SCYA4L* locus polymorphism and association with HIV infection

To evaluate the incidence of *SCYA4L* locus polymorphism in the overall population, we sequenced the junction of intron 2 with exon 3 of the *SCYA4L* locus in healthy individuals. Subjects were classified according to the sequence of acceptor splice site as follows: L1L1 (AG), L2L2 (GG) and L1L2 (double peak -A/G- in

Table I. Poblational distribution of *SCYA4L* polymorphism<sup>a</sup>

Donors	Number of Samples			Genomic Frequency (%)			Allelic Frequency (%)	
	L1L1	L1L2	L2L2	L1L1	L1L2	L2L2	Allele L1	Allele L2
Healthy ( <i>n</i> = 220) <sup>b</sup>	154	59	7	70	27	3	83.4	16.6 <sup>c</sup>
HIV ( <i>n</i> = 175) <sup>b</sup>	90	70	15	51.4	40	8.6	71.4	28.6 <sup>c</sup>
AITD ( <i>n</i> = 30)	21	8	1	70	26.7	3.3	82.7	17.3
IDDM-1 ( <i>n</i> = 30)	22	7	1	73.3	23.4	3.3	85	15
HCV ( <i>n</i> = 80)	60	16	4	75	20	5	85	15

<sup>a</sup> The genomic and allelic frequencies were determined in healthy population, HIV<sup>+</sup> patients and other non-related pathologies.

<sup>b</sup> *p* = 0.0003875,  $\chi^2$  = 15.71.

<sup>c</sup> *p* = 0.00016,  $\chi^2$  = 14.24.

the first nucleotide). Of 220 healthy donors tested, 154 were L1L1 (70%), 59 L1L2 (26.8%) and 7 L2L2 (3.2%). If this polymorphism is considered to be a classical biallelic system, derived allelic frequency would be 83.4 and 16.6% for *SCYA4L1* and *SCYA4L2* alleles, respectively (Table I).

Due to the pivotal role played by CCL4 and its receptor CCR5 in HIV infection we decided to conduct an association study between *SCYA4L* polymorphism and HIV infection. In a group of 175 HIV<sup>+</sup> individuals, 51.4% of them were L1L1 (vs 70% in the controls), 40% were L1L2 (vs 27%), and 8.6% were L2L2 (vs 3%) (*p* = 0.00039,  $\chi^2$  = 15.71). Therefore, the frequency of allele L2 is clearly higher in the group of HIV<sup>+</sup> patients (28.6%) compared with the control group (16.6%) (Table I) (*p* = 0.00016,  $\chi^2$  = 14.24).

To evaluate the association of the *SCYA4L* polymorphism with other polymorphisms involved in the progression of HIV, we analyzed the incidence of CCR5 $\Delta$ 32 (39, 41, 42) and stromal cell-derived factor (SDF)-1 3'A (40) alleles in a random sample of 138 and 155 individuals from our HIV<sup>+</sup> group with CCL4 genotype. (Tables II and III). We did not find any distribution suggesting positive or negative cross-association between the *SCYA4L* polymorphism and the CCR5 $\Delta$ 32 or SDF-1 3'A polymorphisms (Tables II and III). The lack of significant differences between the allelic frequencies of these polymorphisms in our HIV<sup>+</sup> and previously reported groups (39–42) allows us to consider the homogeneity of our group, thus avoiding any unexpected bias in distribution. The allelic frequencies reported in healthy populations (Caucasian and Asian populations) are 74–84% for the wild-type SDF and 16–26% for the SDF 3'A allele (40), and are 86–96% for the wild-type CCR5 and 4–14% for the CCR5 $\Delta$ 32 allele (41, 42). We did not find individuals homozygous for SDF 3'A allele due to their potential resistance to AIDS progression in addition to their low theoretical frequency. Similarly we did not find individuals homozygous for CCR5 $\Delta$ 32 allele due to their potential resistance to HIV infection in addition to their low theoretical frequency.

To explore the possibility that CCL4 forms may be associated with other diseases characterized by a maintained immune response, we evaluated the *SCYA4L* polymorphism in groups of patients with hepatitis C virus infection, type-1 diabetes, and auto-

immune thyroid diseases. No significant differences were found between healthy donors and patients in the genomic and allelic frequencies of both *SCYA4L* locus variants, thus suggesting that the observed association with HIV infection is characteristic of this type of infection.

## Discussion

The CCL4 chemokine is encoded by two paralogous genes, *SCYA4* and *SCYA4L* (23), highly related but with different exonic and intronic sequences. We found that these two genes not only express the variants already described (CCL4 and CCL4L), but also alternatively spliced variants lacking exon 2 (CCL4 $\Delta$ 2 and CCL4L $\Delta$ 2). Moreover, we report a new polymorphism in the *SCYA4L* locus that forms nine new mRNAs. In an activated situation, the contribution of the *SCYA4L* locus to overall CCL4 expression is approximately 50% for the original allelic variant (*SCYA4L1*), and only 12% for the polymorphic allelic variant (*SCYA4L2*). The distribution of this polymorphism in HIV<sup>+</sup> patients shows an association with the *SCYA4L2* allele.

The structural analysis of the protein products of *SCYA4* (CCL4) and *SCYA4L* (CCL4L) revealed the importance for the amino acid at position -4 relative to the fourth conserved cysteine. The amino acid at that position in CCL4 protein (Ser<sub>70</sub>) forms a hydrogen bond with amino acid Thr<sub>67</sub> (located three residues upstream), thus conferring structural stability to the molecule (43). However, the amino acid at that position in the CCL4L protein (Gly<sub>70</sub>) could not form this hydrogen bond to stabilize the loop defined by the  $\beta$ -turn between the second and third strand of the  $\beta$ -sheet. This loop is believed to be essential for the binding of CCL4 to the glycosaminoglycans (GAGs) (44), and it has been suggested that the immobilization of chemokines by GAGs forms stable solid-phase chemokine foci and gradients necessary for directing leukocyte trafficking in vivo to increase their effective local concentration (thus increasing their binding to cell surface receptors), and potentially influence chemokine *t*<sub>1/2</sub> in vivo (45–47). Hence, the destabilization of this loop make difficult CCL4L binding to GAGs and therefore modify their functional features in vivo.

Table II. Distribution of CCR5 $\Delta$ 32 polymorphism<sup>a</sup>

Groups	Genomic Frequency (%)			Allelic Frequency (%)	
	wt/wt	wt/ $\Delta$ 32	$\Delta$ 32/ $\Delta$ 32	Allele wt	Allele $\Delta$ 32
HIV <sup>+</sup> ( <i>n</i> = 138)	88.4	11.6	0	94.2	5.8
HIV <sup>+</sup> L1L1 ( <i>n</i> = 75)	90.7	9.3	0	95.3	4.7
HIV <sup>+</sup> L1L2 ( <i>n</i> = 53)	86.8	13.2	0	93.4	6.6
HIV <sup>+</sup> L2L2 ( <i>n</i> = 10)	90	10	0	95	5

<sup>a</sup> Genomic and allelic frequencies were determined for total HIV<sup>+</sup> individuals and also in each *SCYA4L* genotype-based HIV<sup>+</sup> group (L1L1/L1L2/L2L2).

Table III. Distribution of SDF 3' A polymorphism<sup>a</sup>

Groups	Genomic Frequency (%)			Allelic Frequency (%)	
	wt/wt	wt/3' A	3' A/3' A	Allele wt	Allele 3' A
HIV <sup>+</sup> (n = 155)	71	29	0	85.5	14.5
HIV <sup>+</sup> L1L1 (n = 83)	73.5	26.5	0	86.7	13.2
HIV <sup>+</sup> L1L2 (n = 60)	70	30	0	85	15
HIV <sup>+</sup> L2L2 (n = 12)	58.3	41.7	0	79.2	20.8

<sup>a</sup> Genomic and allelic frequencies were determined for total HIV<sup>+</sup> individuals and also in each SCYA4L genotype-based HIV<sup>+</sup> group (L1L1/L1L2/L2L2).

We found that both *SCYA4* and *SCYA4L* loci not only produce the expected CCL4 and CCL4L mRNA but also produce alternatively spliced mRNAs that lack the second exon, which give rise to the CCL4Δ2 and CCL4LΔ2 variants. These two short (29 aa) proteins only maintain the first two amino acids from the CCL4 and CCL4L proteins and lack three of the four cysteine residues critical for intramolecular disulfide bonding. Therefore, CCL4Δ2 and CCL4LΔ2 may not be structurally considered as chemokines, and despite the difficulty in predicting protein folding, these variants do not seem to be able to bind to CCR5 and thus may have no CCL4 activity. Although CCL3 is closely related to CCL4, we could not detect mRNAs from the two loci (*SCYA3* and *SCYA3L1*) lacking exon 2 (data not shown).

The *SCYA4L* polymorphism described in this paper reveals that the *SCYA4L* locus has two allelic variants: the originally described variant (*SCYA4L*) named *SCYAL1* and the new polymorphic variant *SCYA4L2*. The *SCYA4L2* locus has a base substitution in the acceptor splice site of intron 2 that originates a new complex splicing pattern, including the use of six new acceptor splice sites. We notice that the score of alternative acceptor splice sites used in the *SCYA4L2* allelic variant does not reflect the quantity of each mRNA generated, because the method used only takes into account the universal dinucleotide and their flanking sequences, although we know that other mechanisms may contribute to the efficiency of recognition of splice sites (48–51).

CCL4L2 mRNA, the most abundant mRNA derived from *SCYA4L2*, is characterized by the loss of the first 15 bp in exon 3, being Phe<sub>65</sub>, Gln<sub>66</sub>, Thr<sub>67</sub>, Lys<sub>68</sub>, and Arg<sub>69</sub> (position relative to the initial methionine) the five amino acids deleted in the predicted protein. Critical analysis of the conserved amino acids in CC chemokines show that Phe<sub>65</sub>, Thr<sub>67</sub>, and to a lesser degree Lys<sub>68</sub>, are highly conserved residues in this subfamily. In fact, Phe<sub>65</sub> and Gln<sub>66</sub> residues are the last two residues of the second strand of the CCL4 β-sheet, and Thr<sub>67</sub>, Lys<sub>68</sub> and Arg<sub>69</sub> (next to Gly<sub>70</sub>, which is characteristic of CCL4L) form the β-turn between the second and third strand. Thus, the deletion of these five amino acids would affect the whole monomer structure by disturbing the formation of the core of the molecule, the β-sheet. It is known that CCL4, as well as CCL3 and CCL5, tends to self-associate and, thereby, form homodimers, tetramers, or high molecular mass aggregates in vitro, and possibly in vivo under certain conditions, in a process that involves residues Lys<sub>68</sub> and Arg<sub>69</sub> (52); furthermore, it has identified naturally occurring CCL4/CCL3 heterodimers at physiological concentrations (53). We predict that the deletion has a negative effect on the ability of CCL4L2 to form self-aggregates, or heterodimers with CCL3. Additionally, just as in the case of the single amino acid change between CCL4 and CCL4L proteins (Ser<sub>70</sub> → Gly<sub>70</sub>), we may expect that the GAG binding of CCL4L2 will be seriously affected, if not abrogated.

The folding prediction and the functional features of the other *SCYA4L2*-derived proteins (CCL4L2b, c, d, e, and f and CCL4L2Δ2, L2bΔ2) are difficult to establish. The biological rel-

evance of these proteins is unknown and may be related to their low expression level.

Therefore, CCL4 is a highly variable chemokine since a minimum of 12 variants derive from their two codifying loci. These variants are potential targets for CD26/DPP IV and, similar to the originally described CCL4, may be secreted as naturally truncated forms lacking the two NH<sub>2</sub>-terminal amino acids (29, 30).

Our expression studies show that in activated CD8 T cells (at least in L1L1 individuals), the contribution of the *SCYA4L* locus (*SCYA4L1* allelic variant) is very relevant and accounts for approximately half of CCL4 expression (although in L2L2 individuals the contribution of *SCYA4L* locus is lower, only 12%), in contrast with most previous descriptions, which did not consider the expression of the two different CCL4 loci and conferred the overall data to the role of the *SCYA4* locus. Only recently (54), it was reported that while peripheral blood monocytes predominantly express the *SCYA4* locus, peripheral blood B lymphocytes express both *SCYA4* and *SCYA4L* loci in equivalent amounts; however, this study did not take into account the polymorphism of *SCYA4L* that strongly affects the overall expression of this chemokine.

Concerning the low levels of mRNA derived from *SCYA4L2*, we consider that although the slight differences found between the 5' and 3' non-coding regions of *SCYA4L1* and *SCYAL2* allelic variants (data not shown) are probably not responsible for this low expression, a reduced mRNA stability due to a low efficiency of spliceosomes may be a major determinant. This low spliceosome efficiency may be involved in a situation as complex as that found with the *SCYA4L2* allele (a minimum of nine different mature mRNAs, eight of them produced by using alternative acceptor splice sites). In other examples, the use of cryptic splice sites was found to reduce the amount of mature mRNA generated (55).

The in vivo relevance of the complex situation produced by this *SCYA4L* polymorphism was assessed by the analysis of frequencies in disease. Interestingly, a significant difference was only observed in HIV<sup>+</sup> patients. Since no definitive results were obtained in the study of other clinical relevant data (e.g., number of CD4 T cells, copies of virus. . .) and the preliminary experiments (e.g., in vitro analysis of different genotyping infection ability, blocking capability of each form in HIV-1 infection. . .) are still non-conclusive to suggest any additional pathogenic pathways (data not shown), more extensive studies are undoubtedly needed. Nevertheless, we may hypothesize that the lack of many canonical CCL4L mRNA copies transcribed in L1L2 and L2L2 individuals may reduce the ability of CD8 to protect HIV infection by chemokines, specifically CCL4L. This hypothesis agrees with recent data (56) suggesting a great influence of *SCYA3L1* (CCL3L1) gene-containing segmental duplications on HIV-1/AIDS susceptibility. The strong homology between CCL3 and CCL4 and the genetic linkage between *SCYA3L1* and *SCYA4L* genes and their copy number suggest equivalent mechanisms for both genes, remaining to elucidate which one is the gene mainly involved in this susceptibility.

Following the new classification for chemokine ligands established by Zlotnik and Yoshie (4), the formal gene symbols *SCYA4* and *SCYA4L* are currently used to refer to *CCL4* and *CCL4L*. To systematize the complex nomenclature derived from the high number of alternate designations, we suggest naming the two allelic variants of the *SCYA4L* gene as follows: *SCYA4L1* and *SCYA4L2* (i.e., as used in this article), or *SCYA4L\*1* and *SCYA4L\*2* if we strictly follow the guidelines for human gene nomenclature of HGNC (HUGO Gene Nomenclature Committee). The protein derived from *SCYA4L1* is the one currently named *CCL4L*, but it may alternatively be known as *CCL4L1* to combine it with *CCL4L2* as the major protein derived from *SCYA4L2*. Subsequently, the other proteins derived from *SCYA4L2* may be named *CCL4L2b*, *c*, *d*, *e*, and *f*, whereas the variants lacking exon 2 and derived from *SCYA4* and *SCYA4L* loci may be named like the original protein but adding  $\Delta 2$  (e.g., *CCL4\Delta 2*, *CCL4L\Delta 2*, *CCL4L2\Delta 2*, . . .). To unify this nomenclature with the highly related one for *CCL3* (*SCYA3*, *SCYA3L1*, and *SCYA3L2* currently used to reference *CCL3*, *CCL3L*, and *LD78\gamma* pseudogene), we also suggest following previous *CCL4* nomenclature and the guidelines of HGNC and change the gene symbol for *CCL3L* to *SCYA3L* and *LD78\gamma* to *SCYA3LP*, respectively, since *LD78\gamma* is a 5'-truncated pseudogene derived from the *SCYA3L* locus (20).

Data from our study suggest that variants of chemokines, such as *CCL4*, may be a major feature that determines the final role of these key molecules in immune response, increasing the functional opportunities of these molecules involved in many diseases.

## Acknowledgments

We thank Dr. Maria del Pilar Armengol for help in statistical analysis, Dr. Raúl Castaño for critical reading of the manuscript, Rosa Faner, Marco Fernández, and Pepi Caro for general support in the laboratory; we also thank Drs. Jordi Yagüe and Xavier Forn for kindly providing DNA from healthy donors and samples from hepatitis C virus-positive patients, respectively.

## Disclosures

The authors have no financial conflict of interest.

## References

- Springer, T. A. 1994. Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. *Cell* 76:301.
- Luster, A. D. 1998. Chemokines—chemotactic cytokines that mediate inflammation. *N. Engl. J. Med.* 338:436.
- Baggiolini, M. 1998. Chemokines and leukocyte traffic. *Nature* 392:565.
- Zlotnik, A., and O. Yoshie. 2000. Chemokines: a new classification system and their role in immunity. *Immunity* 12:121.
- Murphy, P. M. 1994. The molecular biology of leukocyte chemoattractant receptors. *Annu. Rev. Immunol.* 12:593.
- Premack, B. A., and T. J. Schall. 1996. Chemokine receptors: gateways to inflammation and infection. *Nat. Med.* 2:1174.
- Zlotnik, A., J. Morales, and J. A. Hedrick. 1999. Recent advances in chemokines and chemokine receptors. *Crit. Rev. Immunol.* 19:1.
- Houshmand, P., and A. Zlotnik. 2003. Therapeutic applications in the chemokine superfamily. *Curr. Opin. Chem. Biol.* 7:457.
- Sierra, M. D., F. Yang, M. Narazaki, O. Salvucci, D. Davis, R. Yarchoan, H. H. Zhang, H. Fales, and G. Tosato. 2004. Differential processing of stromal-derived factor-1 $\alpha$  and  $\beta$  explains functional diversity. *Blood* 103:2452.
- Nibbs, R. J., and G. J. Graham. 2003. CCL27/PESKY: a novel paradigm for chemokine function. *Expert Opin. Biol. Ther.* 3:15.
- Proost, P., I. De Meester, D. Schols, S. Struyf, A. M. Lambeir, A. Wuyts, G. Opendakker, E. De Clercq, S. Scharpe, and J. Van Damme. 1998. Amino-terminal truncation of chemokines by CD26/dipeptidyl-peptidase IV: conversion of RANTES into a potent inhibitor of monocyte chemotaxis and HIV-1-infection. *J. Biol. Chem.* 273:7222.
- Proost, P., S. Struyf, D. Schols, C. Durinx, A. Wuyts, J. P. Lenaerts, E. De Clercq, I. De Meester, and J. Van Damme. 1998. Processing by CD26/dipeptidyl-peptidase IV reduces the chemotactic and anti-HIV-1 activity of stromal-cell-derived factor-1 $\alpha$ . *FEBS Lett.* 432:73.
- Van Damme, J., S. Struyf, A. Wuyts, E. Van Coillie, P. Menten, D. Schols, S. Sozzani, I. De Meester, and P. Proost. 1999. The role of CD26/DPP IV in chemokine processing. *Chem. Immunol.* 72:42.
- Proost, P., P. Menten, S. Struyf, E. Schutyser, I. De Meester, and J. Van Damme. 2000. Cleavage by CD26/dipeptidyl peptidase IV converts the chemokine LD78beta into a most efficient monocyte attractant and CCR1 agonist. *Blood* 96:1674.
- Menten, P., A. Wuyts, and J. Van Damme. 2002. Macrophage inflammatory protein-1. *Cytokine Growth Factor Rev.* 13:455.
- Cocchi, F., A. L. DeVico, A. Garzino-Demo, S. K. Arya, R. C. Gallo, and P. Lusso. 1995. Identification of RANTES, MIP-1 $\alpha$ , and MIP-1 $\beta$  as the major HIV-suppressive factors produced by CD8<sup>+</sup> T cells. *Science* 270:1811.
- Dragic, T., V. Litwin, G. P. Allaway, S. R. Martin, Y. Huang, K. A. Nagashima, C. Cayan, P. J. Maddon, R. A. Koup, J. P. Moore, and W. A. Paxton. 1996. HIV-1 entry into CD4<sup>+</sup> cells is mediated by the chemokine receptor CC-CCR-5. *Nature* 381:667.
- Samson, M., F. Libert, B. J. Doranz, J. Rucker, C. Liesnard, C. M. Farber, S. Saragosti, C. Lapoumeroulie, J. Cognaux, C. Forceille, et al. 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382:722.
- Townsend, J. R., L. F. Barcellos, and R. J. Nibbs. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur. J. Immunol.* 32:3016.
- Hirashima, M., T. Ono, M. Nakao, H. Nishi, A. Kimura, H. Nomiya, F. Hamada, M. C. Yoshida, and K. Shimada. 1992. Nucleotide sequence of the third cytokine LD78 gene and mapping of all three LD78 gene loci to human chromosome 17. *DNA Seq.* 3:203.
- Nakao, M., H. Nomiya, and K. Shimada. 1990. Structures of human genes coding for cytokine LD78 and their expression. *Mol. Cell Biol.* 10:3646.
- Irving, S. G., P. F. Zipfel, J. Balke, O. W. McBride, C. C. Morton, P. R. Burd, U. Siebenlist, and K. Kelly. 1990. Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q. *Nucleic Acids Res.* 18:3261.
- Modi, W. S., J. Bergeron, and M. Sanford. 2001. The human MIP-1 $\beta$  chemokine is encoded by two paralogous genes, ACT-2 and LAG-1. *Immunogenetics* 53:543.
- Nibbs, R. J., J. Yang, N. R. Landau, J. H. Mao, and G. J. Graham. 1999. LD78 $\beta$ , a non-allelic variant of human MIP-1 $\alpha$  (LD78 $\alpha$ ), has enhanced receptor interactions and potent HIV suppressive activity. *J. Biol. Chem.* 274:17478.
- Menten, P., S. Struyf, E. Schutyser, A. Wuyts, E. De Clercq, D. Schols, P. Proost, and J. Van Damme. 1999. The LD78 $\beta$  isoform of MIP-1 $\alpha$  is the most potent CCR5 agonist and HIV-1-inhibiting chemokine. *J. Clin. Invest.* 104:R1.
- Aquaro, S., P. Menten, S. Struyf, P. Proost, J. Van Damme, E. De Clercq, and D. Schols. 2001. The LD78 $\beta$  isoform of MIP-1 $\alpha$  is the most potent CC-chemokine in inhibiting CCR5-dependent human immunodeficiency virus type 1 replication in human macrophages. *J. Virol.* 75:4402.
- Miyakawa, T., K. Obaru, K. Maeda, S. Harada, and H. Mitsuya. 2002. Identification of amino acid residues critical for LD78 $\beta$ , a variant of human macrophage inflammatory protein-1 $\alpha$ , binding to CCR5 and inhibition of R5 human immunodeficiency virus type 1 replication. *J. Biol. Chem.* 277:4649.
- Struyf, S., P. Menten, J. P. Lenaerts, W. Put, A. D'Haese, E. De Clercq, D. Schols, P. Proost, and J. Van Damme. 2001. Diverging binding capacities of natural LD78 $\beta$  isoforms of macrophage inflammatory protein-1 $\alpha$  to the CC chemokine receptors 1, 3, and 5 affect their anti-HIV-1 activity and chemotactic potencies for neutrophils and eosinophils. *Eur. J. Immunol.* 31:2170.
- Guan, E., J. Wang, G. Roderiquez, and M. A. Norcross. 2002. Natural truncation of the chemokine MIP-1 $\beta$ /CCL4 affects receptor specificity but not anti-HIV-1 activity. *J. Biol. Chem.* 277:32348.
- Guan, E., J. Wang, and M. A. Norcross. 2004. Amino-terminal processing of MIP-1 $\beta$ /CCL4 by CD26/dipeptidyl-peptidase IV. *J. Cell Biochem.* 92:53.
- Cocchi, F., A. L. DeVico, R. Yarchoan, R. Redfield, F. Cleghorn, W. A. Blattner, A. Garzino-Demo, S. Colombini-Hatch, D. Margolis, and R. C. Gallo. 2000. Higher macrophage inflammatory protein (MIP)-1 $\alpha$  and MIP-1 $\beta$  levels from CD8<sup>+</sup> T cells are associated with asymptomatic HIV-1 infection. *Proc. Natl. Acad. Sci. USA* 97:13812.
- Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268:78.
- Krogh, A. 1997. Two methods for improving performance of an HMM and their application for gene finding. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 5:179.
- Reese, M. G., F. H. Eeckman, D. Kulp, and D. Haussler. 1997. Improved splice site detection in Genie. *J. Comput. Biol.* 4:311.
- Shapiro, M. B., and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15:7155.
- Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon. 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA* 75:4853.
- Breathnach, R., and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349.
- Adreani, P., I. Gomez, R. Colobran, P. Caro, F. Pelusa, R. Pujol-Borrell, and M. Juan. 2001. Semiquantitative assessment of MIP-1 $\alpha$  and MIP-1 $\beta$  mRNA isoform expression by using a recombinant competitor fragment and RT-AFLP. *Immunologia* 20:184.
- Liu, R., W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup, and N. R. Landau. 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply exposed individuals to HIV-1 infection. *Cell* 86:367.
- Winkler, C., W. Modi, M. W. Smith, G. W. Nelson, X. Wu, M. Carrington, M. Dean, T. Honjo, K. Tashiro, D. Yabe, et al. 1998. Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant: ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study



- (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). *Science* 279:389.
41. Libert, F., P. Cochaux, G. Beckman, M. Samson, M. Aksenova, A. Cao, A. Czeizel, M. Claustres, C. de la Rúa, M. Ferrari, et al. 1998. The  $\delta$ cr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum. Mol. Genet.* 7:399.
  42. Martinson, J. J., N. H. Chapman, D. C. Rees, Y. T. Liu, and J. B. Clegg. 1997. Global distribution of the CCR5 gene 32-basepair deletion. *Nat. Genet.* 16:100.
  43. Lodi, P. J., D. S. Garrett, J. Kuszewski, M. L. Tsang, J. A. Weatherbee, W. J. Leonard, A. M. Gronenborn, and G. M. Clore. 1994. High-resolution solution structure of the  $\beta$  chemokine hMIP-1 $\beta$  by multidimensional NMR. *Science* 263:1762.
  44. Koopmann, W., C. Ediriwickrema, and M. S. Krangel. 1999. Structure and function of the glycosaminoglycan binding site of chemokine macrophage-inflammatory protein-1 $\beta$ . *J. Immunol.* 163:2120.
  45. Tanaka, Y., D. H. Adams, S. Hubscher, H. Hirano, U. Siebenlist, and S. Shaw. 1993. T-cell adhesion induced by proteoglycan-immobilized cytokine MIP-1 $\beta$ . *Nature* 361:79.
  46. Ali, S., A. C. Palmer, B. Banerjee, S. J. Fritchley, and J. A. Kirby. 2000. Examination of the function of RANTES, MIP-1 $\alpha$ , and MIP-1 $\beta$  following interaction with heparin-like glycosaminoglycans. *J. Biol. Chem.* 275:11721.
  47. Proudfoot, A. E., T. M. Handel, Z. Johnson, E. K. Lau, P. LiWang, I. Clark-Lewis, F. Borlat, T. N. Wells, and M. H. Kosco-Vilbois. 2003. Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. *Proc. Natl. Acad. Sci. USA* 100:1885.
  48. Roca, X., R. Sachidanandam, and A. R. Krainer. 2003. Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res.* 31:6321.
  49. Miriami, E., H. Margalit, and R. Sperling. 2003. Conserved sequence elements associated with exon skipping. *Nucleic Acids Res.* 31:1974.
  50. Brudno, M., M. S. Gelfand, S. Spengler, M. Zorn, I. Dubchak, and J. G. Conboy. 2001. Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res.* 29:2338.
  51. Harris, N. L., and P. Senapathy. 1990. Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res.* 18:3015.
  52. Czaplowski, L. G., J. McKeating, C. J. Craven, L. D. Higgins, V. Appay, A. Brown, T. Dudgeon, L. A. Howard, T. Meyers, J. Owen, et al. 1999. Identification of amino acid residues critical for aggregation of human CC chemokines macrophage inflammatory protein (MIP)-1 $\alpha$ , MIP-1 $\beta$ , and RANTES. Characterization of active disaggregated chemokine variants. *J. Biol. Chem.* 274:16077.
  53. Guan, E., J. Wang, and M. A. Norcross. 2001. Identification of human macrophage inflammatory proteins 1 $\alpha$  and 1 $\beta$  as a native secreted heterodimer. *J. Biol. Chem.* 276:12404.
  54. Lu, J., M. Honczarenko, and S. R. Sloan. 2004. Independent expression of the two paralogous CCL4 genes in monocytes and B lymphocytes. *Immunogenetics* 55:706.
  55. Vreken, P., R. W. Niessen, M. Peters, M. C. Schaap, J. G. Zuihoff-Rijntjes, and A. Sturk. 1995. A point mutation in an invariant splice acceptor site results in a decreased mRNA level in a patient with severe coagulation factor XIII subunit A deficiency. *Thromb. Haemost.* 74:584.
  56. Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R. J. Nibbs, B. I. Freedman, M. P. Quinones, M. J. Bamshad, K. K. Murthy, et al. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307:1434.

## 4.2. CAPÍTOL II

### OBJECTIU:

Definir quines són les confusions més habituals i rellevants en les seqüències de CCL4 i CCL4L presents en els catàlegs d'algunes empreses subministradores: clarificar l'origen de les seqüències i remarcar les diferències de cada variant per tal d'evitar l'ús de productes erronis

### ARTICLE:

**Confusión entre CCL4 y CCL4L1: Un ejemplo a tener en cuenta cuando se usan reactivos de terceras partes.**

R. Colobran, M. Juan.

*Inmunología*, Vol. 26 / Núm 1/ Enero-Marzo **2007**: 51-54.



#### 4.2.1. CONSIDERACIONS PRÈVIAS SOBRE L'ARTICLE

Tal i com s'ha comentat en l'apartat anterior, un dels objectius inicials del projecte era realitzar estudis funcionals per a detectar possibles diferències entre CCL4 i CCL4L. El complex panorama genètic descrit pel tàndem *CCL4-CCL4L* va posposar aquest objectiu i, de fet, vam considerar oportú afegir algunes de les noves variants als futurs estudis funcionals. Per fer una primera aproximació, però, es va plantejar la possibilitat d'adquirir les proteïnes CCL4 i CCL4L procedents d'alguna de les diverses empreses que les comercialitzen.

#### 4.2.2. RESUM DE DADES DEL SEGON CAPÍTOL:

Tenint en compte la manca d'estudis amb la proteïna CCL4L i la confusa nomenclatura tradicional entre CCL4 i CCL4L, alhora de fer la búsqueda entre els diferents subministradors possibles, no ens vam limitar a comprovar que la descripció del producte fos l'adequada, sinó que vam accedir a la seqüència aminoacídica per tal de contrastar-la. La nostra sorpresa va ser que, en alguns casos, les seqüències adscrites a CCL4 i CCL4L, tenien errors importants que invalidaven totalment l'ús d'aquestes proteïnes en qualsevol estudi mínimament rigorós. Cal destacar que entre les empreses que presentaven errors de seqüència n'hi havia de punteres en la comercialització de proteïnes recombinants i, concretament, de citocines i quimiocines. Malgrat que de seguida vam advertir-los d'aquestes circumstàncies, la no resolució total d'aquestes indefinicions va fer que consideréssim la possibilitat de posar en coneixement aquests fets a la comunitat científica. L'objectiu era evitar que els investigadors arribessin a conclusions imprecises en base a uns resultats obtinguts amb uns productes erronis, alhora que preteníem alertar-los de la necessitat de contrastar a fons les característiques dels productes que ens arriben de terceres parts (com són les empreses comercials). Cal dir, però, que l'article que es presenta a continuació no pretén qüestionar la qualitat general dels productes de les empreses citades (que, normalment, és força elevada). Per acabar es mostra el full de dades de la proteïna recombinant CCL4L1 que ofereix una altra de les grans empreses del sector com a bon exemple d'informació d'un producte. A la descripció del producte es citen els articles que proporcionen les dades essencials pel consumidor. En aquest cas concret, a la descripció de la quimiocina CCL4L1 apareix l'article que forma part del capítol I d'aquesta tesi doctoral [123].

Finalment, tant sols afegir que, de cara a l'objectiu de realitzar estudis funcionals amb les diferents variants de CCL4 i CCL4L, hem optat per sintetitzar totes les proteïnes d'interès de forma recombinant. Aquest objectiu no forma part d'aquesta tesi doctoral, sinó que n'és la continuació, i s'està desenvolupant actualment al nostre laboratori.



## Confusión entre CCL4 y CCL4L1: Un ejemplo a tener en cuenta cuando se usan reactivos de terceras partes

R. Colobran, M. Juan

*Laboratori d'Immunobiologia per a la Recerca i Aplicacions Diagnòstiques (LIRAD), Banc de Sang i Teixits (BST), Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IICSGTIP). Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona (UAB).*

### CONFUSION BETWEEN CCL4 AND CCL4L1: AN EXAMPLE TO BEAR IN MIND WHEN USING THIRD-PARTY REAGENTS

Recibido: 6 Febrero 2007

Aceptado: 13 Marzo 2007

#### RESUMEN

¿Es siempre fiable un producto comercial? Utilizando como ejemplo las quimiocinas CCL4 y CCL4L1, este trabajo muestra cómo existen cambios trascendentes en las secuencias de un producto comercial. Las confusiones detectadas en las secuencias de CCL4 y CCL4L1 pueden conllevar consecuencias importantes, especialmente cuando se pretenden definir actividades funcionales o efectos celulares de dichas moléculas. Remarcando estos errores queremos también alertar a los investigadores de su responsabilidad al valorar todos los datos existentes sobre una molécula concreta, basados en resultados obtenidos a partir de un producto comercial, pudiendo llevar a cuestionar los resultados que con su uso experimental se puedan desprender. Estas diferencias son siempre importantes, pero lo son más cuando afectan a proveedores mayoritarios del sector, puesto que las interpretaciones erróneas que se desprenden pueden irse perpetuando a lo largo de la investigación de ese campo.

PALABRAS CLAVE: Quimiocinas / Control de calidad.

#### ABSTRACT

Is a commercial product always reliable? Using CCL4 and CCL4L1 chemokines as a model, this work shows important changes in the sequences of a commercial product. The detected confusions in CCL4 and CCL4L1 can entail important consequences, particularly in experiments aimed at defining functional activities or cellular effects of these molecules. We point to these errors to alert researchers of their responsibility in considering all the existing data on a specific molecule obtained from a commercial source, even leading to question the results that could be obtained by their experimental use. These sequence differences are always important, but even more so when major suppliers are involved, since wrong interpretations can be perpetuated in a particular research field.

KEY WORDS: Chemokines / Quality Control.

Dentro de la superfamilia de las quimiocinas y, concretamente, en la subfamilia CC encontramos CCL3/MIP-1 $\alpha$  y CCL4/MIP-1 $\beta$ , dos quimiocinas inflamatorias altamente relacionadas<sup>(1)</sup>. CCL3 y CCL4 han evolucionado por duplicación desde un gen ancestral común y ambas quimiocinas presentan una segunda copia no alélica (CCL3L1/LD78 $\beta$  y CCL4L1/LAG-1) presente en número variable en el genoma humano<sup>(2)</sup>. Estas dos quimiocinas comparten la unión a CCR5 que, a su vez, es correceptor de las cepas M-trópicas del virus VIH-1.

Recientemente, el estudio inmunogenético de CCL4 y CCL4L1 ha aportado nuevos datos que aumentan la complejidad de estas dos moléculas mediante fenómenos de corte y unión (*splicing*) alternativo del ARN, y por la presencia de un polimorfismo que afecta el patrón de transcritos de CCL4L1, generando nuevas variantes a partir de estos dos genes<sup>(3)</sup>.

Como para otras quimiocinas, actualmente se pueden adquirir comercialmente tanto CCL4 como CCL4L1 producidas de forma recombinante. Aunque desde el año 2000 existe una nomenclatura unificadora para todas las quimiocinas y sus receptores<sup>(4)</sup>, algunas casas comerciales aún utilizan la denominación antigua aunque la situación más frecuente es que utilicen el nombre o nombres clásicos junto al nuevo (por ejemplo CCL4/MIP-1 $\beta$ /ACT-2).

Pero toda esta complejidad supone que cada una de estas variantes debe estar bien definida y, aunque puede parecer sorprendente, hemos observado que se producen indefiniciones en los productos ofrecidos por algunas empresas. De hecho, el motivo de este escrito es alertar de estas posibles indefiniciones a los usuarios y compradores potenciales de quimiocinas, especialmente de aquellas que, como CCL4 y CCL4L1, tienen variantes altamente homólogas.

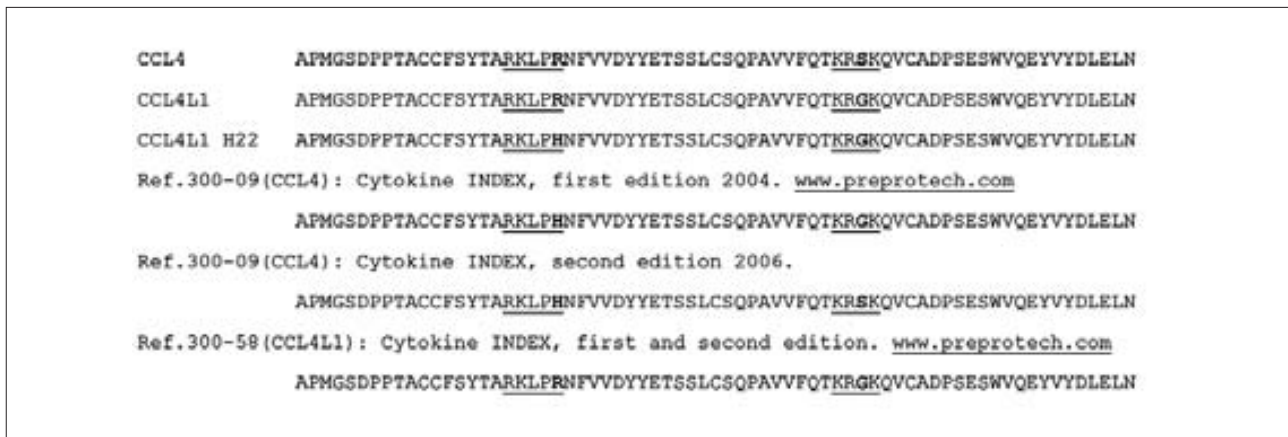
El caso que aquí presentamos se refiere a la confusión existente en cuanto a CCL4 y CCL4L1. Más allá de ser útil para aquellos que estudian o usan CCL4, este trabajo plantea, partiendo de la evidencia concreta de los datos de CCL4 y CCL4L1, la idea general de que es totalmente necesario contrastar todas las características de un producto que se obtiene de terceras partes (habitualmente empresas que lo comercializan) y que se usa de manera central en un estudio experimental. Ello es especialmente relevante ante las reticencias de muchos proveedores para proporcionar datos complementarios de los reactivos, excusándose en la protección de su producción comercial. Sin una evaluación rigurosa de estos datos, es posible llegar a "demostrar" conceptos erróneos, que, especialmente cuando los proveedores son mayoritarios en el campo, pueden llegar a perpetuarse como ciertos a lo largo de trabajos sucesivos.

Este hecho ha sido observado de forma muy evidente en el caso de Peprotech (London, UK), una empresa que suministra, a nivel mundial, productos relacionados con el mundo de las citocinas. Más allá de la aceptada, y mayoritariamente contrastada, calidad de los productos de Peprotech (que es una de las empresas líderes en el sector), en su catálogo "Cytokine INDEX" encontramos las quimiocinas CCL4/MIP-1 $\beta$  y CCL4L1/LAG-1 con confusiones importantes en sus secuencias aminoacídicas definidas en este catálogo. De hecho, los cambios más importantes se detectaron en la primera edición de este catálogo (año 2004). Nuestro laboratorio puso en conocimiento de la empresa dichos cambios (2006), y más allá de agradecer nuestra precisión, no recibimos respuesta que justificase o pudiese al menos cuestionar las consecuencias de estos errores. De hecho, en la segunda edición del "Cytokine INDEX" del año 2006 (y que es la que en el momento de redacción de este trabajo está vigente) persisten aún algunos de ellos (aunque otros se han subsanado), sin que en ningún sitio se justifiquen o se comenten las razones de estos cambios. Seguidamente se describen las confusiones en los dos catálogos:

#### CYTOKINE INDEX, FIRST EDITION 2004

En el caso de CCL4/MIP-1 $\beta$  (ref. 300-09) encontramos que la secuencia presente (tanto en su catálogo escrito como en su página de Internet) contiene dos aminoácidos incorrectos: En la posición 22 aparece una histidina (H) en vez de una arginina (R). Este error puede deberse a la existencia de un polimorfismo descrito en esta posición (rs3744595), pero en su homóloga CCL4L1<sup>(5)</sup> y no en CCL4 (según datos propios sobre 81 individuos de nuestro país, este polimorfismo tiene una frecuencia alélica poblacional en nuestra región del 8,05% (manuscrito en preparación)). Además, en la posición 47 hay una glicina (G) en vez de una serina (S). Este cambio aminoacídico es precisamente la única diferencia que hay entre CCL4 (S<sub>47</sub>) y CCL4L1 (G<sub>47</sub>). Por tanto, bajo el nombre de CCL4 encontramos una secuencia que corresponde a una variante polimórfica poco frecuente de CCL4L1 (CCL4L1 H22) (Fig. 1).

En el caso de la quimiocina CCL4L1/LAG-1 (ref. 300-58) existe, tanto en su catálogo escrito como en su página de Internet, la siguiente descripción: "LAG-1 is identical to MIP-1 $\beta$  (ACT II isotype) except for one amino acid substitution of arginine for histidine at position 22 of the mature protein". Esta descripción es incorrecta ya que la única diferencia entre CCL4/MIP-1 $\beta$  y CCL4L1/LAG-1 a nivel de proteína madura es el cambio ya mencionado Ser<sub>47</sub> (CCL4)  $\rightarrow$  Gly<sub>47</sub> (CCL4L1). Además la sustitución R22H corresponde, como se ha dicho anteriormente, a una variante polimórfica poco frecuente de la propia CCL4L1. Sorprendentemente,



**Figura 1.** Alineamiento de las secuencias de CCL4, CCL4L1 y CCL4L1 H22, y de las referencias que figuran en los catálogos de Peprotech como CCL4 (ref. 300-09) y CCL4L1 (ref. 300-58). En negrita se destacan los aminoácidos de la posición 22 (R/H) y 47 (S/G). Los motivos BBXXB y BBXB implicados en la unión a GAGs están subrayados.

contradiendo la propia descripción, la secuencia que se muestra es la realmente correcta para CCL4L1 (con los aminoácidos R<sub>22</sub> y G<sub>47</sub>) (Fig. 1).

**CYTOKINE INDEX, SECOND EDITION 2006**

En el caso de CCL4/MIP-1β (ref. 300-09) se mantiene el error en el aminoácido 22 [aparece una histidina (H) en vez de una arginina (R)]. Por otra parte se ha corregido la confusión en la posición 47, donde ahora aparece la serina (S) que define esta quimiocina. Sorprendentemente esta corrección no aparece en su página de Internet a día de hoy (febrero 2007).

En el caso de CCL4L1/LAG-1 (ref. 300-58) no hay cambios respecto la edición del 2004 (se mantiene la descripción incorrecta tanto en el catálogo como en Internet pero la secuencia es correcta).

Las consecuencias de los errores que aquí se explicitan pueden variar en función de las aplicaciones para las que se adquieran estas quimiocinas. Existen muy pocos datos sobre la diversidad funcional entre CCL4 y CCL4L1 ya que la descripción genética es de 1990<sup>(6)</sup> pero la constatación de la importancia de diferenciar entre ambas quimiocinas se hace patente hace tan sólo unos tres años<sup>(7)</sup>. Aunque estos datos no permiten definir diferencias significativas a nivel de quimiotaxis, afinidad de unión a CCR5 y bloqueo de la infección por VIH<sup>(8)</sup>, diferencias funcionales entre CCL4 y CCL4L1 son esperables y en ningún caso deben excluirse, puesto que hasta ahora todas las evidencias funcionales descritas han sido obtenidas siempre en ensayos *in vitro*. Así son esperables funciones distintas, explicables por el hecho de que el único aminoácido de diferencia entre

CCL4 y CCL4L1 (Ser<sub>47</sub>→Gly<sub>47</sub>) se encuentra formando parte del motivo BBXB (donde B representa un aminoácido básico y X equivale a la Ser<sub>47</sub> o Gly<sub>47</sub>), crítico para la unión de las CC quimiocinas inflamatorias a los glicosaminoglicanos (GAGs)<sup>(9, 10)</sup>. Dicha unión se ha descrito recientemente como esencial para la funcionalidad *in vivo* de ciertas quimiocinas (entre ellas CCL4)<sup>(11)</sup>, y por tanto se deben prever diferencias entre CCL4 y CCL4L1. Por otra parte la sustitución R22H (que corresponde a una variante polimórfica de CCL4L1) también debiera tener influencia en el mismo sentido, ya que la arginina 22 forma parte del motivo BBXXB que, junto al ya mencionado BBXB, forman las dos regiones de CCL4 implicadas en la unión a los GAGs (Fig. 1). Hasta el momento existe una única publicación que pone en relieve la importancia del cambio R22H, indicando que los individuos infectados por el virus VIH y con la variante CCL4L1 H22 (ya sea en homocigosis o heterocigosis) exhiben una peor supervivencia en comparación a aquellos que tienen la variante común CCL4L1 R22 en homocigosis<sup>(5)</sup>.

Cabe mencionar que la misma secuencia “errónea” que Peprotech presenta para CCL4/MIP-1β también es ofrecida por Research Diagnostics, Inc. (ref. RDI-309). En cambio esta casa ofrece también CCL4L1/LAG-1 con la secuencia correcta (ref. RDI-30589). Hay que decir que existen bastantes otras casas comerciales que ofrecen CCL4/MIP-1β y unas pocas que también ofrecen CCL4L1/LAG-1. Muchas de ellas no concretan la secuencia aminoacídica en su catálogo o página de Internet. De las casas comerciales que ofrecen estas quimiocinas, y en las que sí se puede comprobar las secuencias aminoacídicas, destacamos las siguientes: R&D ofrece CCL4/MIP-1β (ref. 271-BME), USBiological ofrece CCL4L1/LAG-1 (ref. L1050), ProSpec-Tany TechnoGene



LTD ofrece CCL4/MIP-1 $\beta$  (ref. CHM-276) y Bioclone Inc. ofrece CCL4/MIP-1 $\beta$  (ref. RCB-009).

Definir en qué grado estos cambios pueden haber inducido errores de interpretación en los resultados publicados hasta la fecha, es tarea difícil de realizar cuando la bibliografía sobre CCL4 (MIP-1 $\beta$ ) contiene más de 2.000 artículos y, aunque no todos involucran el uso de proteína recombinante humana, son muchos los trabajos que usan estos productos en la valoración de la función de la quimiocina. La falta de herramientas bioinformáticas para seleccionar los artículos en base a la presencia de determinados productos en el apartado de material y métodos hace muy complejo poder valorar la repercusión de los problemas apuntados.

En resumen, estamos ante una situación que ejemplifica la problemática que supone confiar trabajos experimentales en base a productos de terceras partes. Creemos que es responsabilidad de los investigadores valorar con rigurosidad los datos existentes de una molécula concreta, basados en resultados obtenidos a partir de un producto comercial, especialmente cuando se pretenden definir actividades funcionales o efectos celulares en moléculas producidas por terceras partes. Esto es especialmente difícil cuando los proveedores, por supuestos "motivos de seguridad en la producción", no aportan datos relevantes de los productos que venden. Las diferencias son importantes, especialmente cuando afectan a proveedores mayoritarios del sector, puesto que los errores de interpretación pueden irse perpetuando entre investigadores. Parece claro que los investigadores independientes deberían siempre contrastar toda la información disponible del producto que obtienen y exigirla a los proveedores de los productos utilizados cuando los resultados sean novedosos o supongan incongruencias con resultados previos. Este criterio también debería aplicarse en las revisiones de los artículos, exigiendo los datos pertinentes a esta información básica. Los autores de este trabajo consideramos especialmente relevante que, en el caso de productos recombinantes, se disponga y evalúen los datos de secuencia y de método de producción.

#### AGRADECIMIENTOS

Este trabajo ha sido posible gracias a las ayudas del Fondo de Investigaciones Sanitarias (proyecto 02/0104) y la Fundación para la Investigación y la Prevención del SIDA en España (FIPSE, proyecto 36487/05).

#### CONFLICTO DE INTERÉS

Los autores declaran no tener conflicto de interés.

#### CORRESPONDENCE TO:

Dr. Manel Juan  
Laboratori d'Immunobiologia per a la Recerca i Aplicacions  
Diagnòstiques (LIRAD)  
Banc de Sang i Teixits (BST)  
Edifici Escoles. Ctra de Can Ruti s/n  
08916 Badalona (Barcelona), Spain  
Phone number: 34-93.497.86.66. Fax: 34-93.497.86.68  
E-mail: mjuan.liradbst.germanstria@gencat.net

#### REFERENCES

1. Menten P, Wuyts A, Van Damme J. Macrophage inflammatory protein-1. *Cytokine Growth Factor Rev* 2002;13:455-481.
2. Townson JR, Barcellos LF, Nibbs RJ. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol* 2002;32:3016-26.
3. Colobran R, Adreani P, Ashhab Y, et al. Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. *J Immunol* 2005;174:5655-5664.
4. Zlotnik A, Yoshie O. Chemokines: a new classification system and their role in immunity. *Immunity* 2000;12:121-127.
5. Capoulade-Metay C, Meyer L, Tran T, et al. Influence of the R22H variant of macrophage inflammatory protein 1, /Lag-1 in HIV-1 survival. *Aids* 2005;19:831-833.
6. Baixeras E, Roman-Roman S, Jitsukawa S, Genevee C, Mechiche S, Viegas-Pequignot E, Hercend T, Triebel F. Cloning and expression of a lymphocyte activation gene (LAG-1). *Mol Immunol*. 1990;27:1091-1102.
7. Modi WS, Bergeron J, Sanford M. The human MIP-1, chemokine is encoded by two paralogous genes, ACT-2 and LAG-1. *Immunogenetics*. 2001;53:543-549.
8. Howard OM, Turpin JA, Goldman R, Modi WS. Functional redundancy of the human CCL4 and CCL4L1 chemokine genes. *Biochem Biophys Res Commun* 2004;320:927-931.
9. Koopmann W, Ediriwickrema C, Krangel MS. Structure and function of the glycosaminoglycan binding site of chemokine macrophage-inflammatory protein-1 beta. *J Immunol* 1999;163:2120-2127.
10. Ali S, Palmer AC, Banerjee B, Fritchley SJ, Kirby JA. Examination of the function of RANTES, MIP-1 $\alpha$ , and MIP-1 $\beta$  following interaction with heparin-like glycosaminoglycans. *J Biol Chem* 2000;275:11721-11727.
11. Proudfoot AE, Handel TM, Johnson Z, Law EK, LiWang P, Clark-Lewis I, et al. Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. *Proc Natl Acad Sci USA* 2003;100:1885-1890.

### Specifications and Use

- |                        |   |
|------------------------|---|
| <b>Source</b>          | ◆ A DNA sequence encoding the mature human CCL4L1 (Ala 24 - Asn 92) (Accession # NP_001001435; Zipfel, P.F. <i>et al.</i> , 1989, J. Immunol. <b>142</b> (5):1582 - 1590) was expressed in <i>E. coli</i> .   |
| <b>Molecular Mass</b>  | ◆ Based on N-terminal sequencing, the recombinant human CCL4L1 starts at Ala 24 and has a calculated molecular mass of approximately 7.8 kDa. Recombinant human CCL4L1 migrates as an approximately 12 kDa protein in SDS-PAGE under reducing conditions.   |
| <b>Purity</b>          | ◆ > 95%, as determined by SDS-PAGE and visualized by silver stain.  |
| <b>Endotoxin Level</b> | ◆ < 1.0 EU per 1 $\mu$ g of the cytokine as determined by the LAL method.   |
| <b>Activity</b>        | ◆ Measured by its ability to chemoattract mouse BaF/3 cells transfected with human CCR5.<br>◆ The ED <sub>50</sub> for this effect is typically 0.5 - 2.5 ng/mL.  |
| <b>Formulation</b>     | ◆ Lyophilized from a 0.2 $\mu$ m filtered solution in PBS containing 50 $\mu$ g of bovine serum albumin per 1 $\mu$ g of cytokine.  |
| <b>Reconstitution</b>  | ◆ It is recommended that sterile PBS containing at least 0.1% human serum albumin or bovine serum albumin be added to the vial to prepare a stock solution of no less than 10 $\mu$ g/mL.   |
| <b>Storage</b>         | ◆ Lyophilized samples are stable for up to twelve months from date of receipt at -20° C to -70° C.<br>◆ Upon reconstitution, this cytokine can be stored under sterile conditions at 2° - 8° C for one month or at -20° C to -70° C <b>in a manual defrost freezer</b> for three months without detectable loss of activity.<br>◆ <b>Avoid repeated freeze-thaw cycles.</b> |

### Human CCL4L1

CCL4, also known as MIP-1 $\beta$ , is an 12 kDa  $\beta$  chemokine that is secreted by activated leukocytes, lymphocytes, vascular endothelial cells, and pulmonary smooth muscle cells. CCL4 interacts with CCR5 to attract lymphocytes, NK cells, and immature dendritic cells to sites of inflammation.<sup>1-4</sup> CCL4 also blocks the entry of HIV into CCR5-expressing cells.<sup>5</sup> Human CCL4 and its variants are encoded by two paralogous genes,<sup>6-8</sup> one of which (SCYA4) encodes CCL4 itself.<sup>9</sup> The other gene (SCYA4L) is found in two alleles that give rise to CCL4L1 and CCL4L2, respectively.<sup>9</sup> The human CCL4L1 cDNA encodes a 92 amino acid (aa) precursor with a 23 aa signal sequence. Alternate splicing yields an isoform with a 40 aa internal deletion. The second allele (SCYA4L2), encoding CCL4L2, is spliced into a variety of isoforms.<sup>9</sup> Human CCL4L1 shares greater than 98% aa sequence identity with CCL4 and CCL4L2. It shares 96% aa sequence identity with rhesus CCL4 and approximately 80 - 90% aa sequence identity with bovine, mouse, rabbit, and rat CCL4. The gene copy number for CCL4L1 varies from zero to five, but this is not associated with changes in mRNA or protein levels.<sup>10</sup> CCL4 and CCL4L1 demonstrate comparable CCR5 binding, promotion of chemotaxis, and inhibition of HIV replication in stimulated PBMC.<sup>10,11</sup> *In vivo*, the first two N-terminal amino acids of CCL4 are removed by DPPIV. The truncated form is as active as full-length CCL4 and gains the ability to bind CCR1 and CCR2b.<sup>12,13</sup> Similar proteolytic processing of CCL4L1 has not been described.

#### References:

1. Maurer, M. and E. von Stebut, 2004, Int. J. Biochem. Cell Biol. **36**:1882.
2. Menten, P. *et al.*, 2002, Cytokine Growth Factor Rev. **13**:455.
3. Bisset, L.R. and P. Schmid-Grendelmeier, 2005, Curr. Opin. Pulm. Med. **11**:35.
4. Frangogiannis, N.G., 2004, Inflamm. Res. **53**:585.
5. Lehner, T., 2002, Trends Immunol. **23**:347.
6. Modi, W.S., 2004, Genomics **83**:735.
7. Lu, J. *et al.*, 2004, Immunogenetics **55**:706.
8. Modi, W.S., 2001, Immunogenetics **53**:543.
9. Colobran, R. *et al.*, 2005, J. Immunol. **174**:5655.
10. Townson, J.R. *et al.*, 2002, Eur. J. Immunol. **32**:3016.
11. Howard, O.M.Z. *et al.*, 2004, Biochem. Biophys. Res. Commun. **320**:927.
12. Guan, E. *et al.*, 2004, J. Cell. Biochem. **92**:53.
13. Guan, E. *et al.*, 2002, J. Biol. Chem. **277**:32348.

FOR RESEARCH USE ONLY. NOT FOR USE IN HUMANS.

**R&D Systems, Inc.**  
**1-800-343-7475**



## 4.3. CAPÍTOL III

### OBJECTIU:

Determinar la distribució poblacional mundial del nombre de còpies de *CCL4L* i dels seus SNPs més rellevants, utilitzant un protocol que ens permeti discriminar el nombre exacte de còpies de cada al·lel en individus heterozigots amb més de 2 còpies de *CCL4L*. Establir si existeix correlació entre el nombre de còpies de *CCL4L* i les freqüències dels SNPs.

### ARTICLE:

#### **Population Structure in Copy Number Variation (CNV) and SNPs in the CCL4L Gene.**

Roger Colobran, David Comas, Rosa Faner, Edurne Pedrosa, Roger Anglada, Ricardo Pujol-Borrell, Jaume Bertranpetit, Manel Juan.

*Genes and Immunity*, Submitted.



### 4.3.1. CONSIDERACIONS PRÈVIES SOBRE L'ARTICLE

Durant la fase final de l'estudi descrit al capítol I, va sortir publicat un treball que evidenciava que els gens *CCL3L* i *CCL4L* es trobaven en nombre variable de còpies en el genoma humà [107]. Aquest estudi, realitzat en població caucàsica, també demostrava que, en el cas de *CCL3L*, el nombre de còpies del gen estava relacionat amb els nivells d'expressió de la proteïna. Curiosament, no trobaven aquesta relació en el cas de *CCL4L*. Malgrat que les raons que expliquen aquesta manca de correlació per a *CCL4L* poden ser diverses cal destacar que: a) aquest article presentava incongruències de seqüència en la definició de *CCL4L* respecte *CCL4* i, b) els autors desconeixien l'existència de les dues variants al·lèliques principals de *CCL4L* (*CCL4L1* i *CCL4L2*), publicades posteriorment a aquest treball, i que poden ser importants ja que comporten diferències notables d'expressió.

### 4.3.2. RESUM DE DADES DEL TERCER CAPÍTOL:

La rellevància del fet que *CCL4L* (a l'igual que *CCL3L*) es trobés en nombre variable de còpies al genoma humà, va fer plantejar-nos l'aprofundir sobre aquest tema. A més a més, tenint en compte la detallada descripció del complex panorama genètic i transcripcional de *CCL4-CCL4L* que havíem realitzat [123], no podíem deixar d'estudiar un aspecte tan important com la CNV que afectava a *CCL4L*. Amb aquest propòsit es va dissenyar un protocol per quantificar el nombre de còpies gèniques de *CCL4L* per PCR a temps real, inicialment utilitzant sondes FRET (utilitzant un aparell Lightcycler<sup>®</sup>, de Roche) i posteriorment adaptant-lo a l'ús de sondes TaqMan<sup>®</sup> (per utilitzar als termocicladors a temps real d'Applied Biosystems). En ambdós casos, la idea bàsica del mètode és comparar en una mateixa mostra (sistema *multiplex*) l'amplificació del gen *CCL4L* amb un gen de referència que presenti sempre 2 còpies/genoma diploide.

Per a tenir una idea de l'estructura i distribució poblacional del nombre de còpies de *CCL4L*, vam decidir que seria interessant incloure mostres de poblacions humanes de diferents parts del planeta. La col·laboració amb els doctors Jaume Bertranpetit i David Comas del departament de biologia evolutiva de la Universitat Pompeu Fabra, ens va permetre utilitzar per a aquest estudi el panell de mostres HGDP-CEPH [sigles corresponents al *Human Genome Diversity Project* (HGDP) de la *Foundation Jean Dausset* (CEPH) a París] [124]. Bàsicament es tracta de 1064 mostres de DNA genòmic (gDNA) corresponents a 1051 individus de 51 poblacions mundials diferents. Aquestes mostres van ser recollides sota criteris molt estrictes per diversos laboratoris pertanyents al HGDP i al CEPH i la informació de cada mostra es limita al sexe de l'individu i la població i origen geogràfic. El panell de mostres HGDP-CEPH ha estat i és utilitzat en nombrosos estudis genètics de prestigi [109, 125-127]. El fet d'haver d'analitzar un nombre tan

elevat de mostres va fer decidir-nos per l'ús del protocol amb sondes TaqMan per la possibilitat d'utilitzar plaques de 384 pous en un termociclador 7900 d'Applied Biosystems.

Paral·lelament a la quantificació del nombre de còpies de *CCL4L*, preteníem també genotipar les mateixes mostres per a l'SNP descrit en l'article del capítol I (rs4796195) i per un altre SNP de *CCL4L* (rs3744595) que s'havia publicat feia poc com un factor d'influència en la supervivència dels individus infectats pel virus HIV [128]. Un dels problemes *a priori* de la genotipificació era (i és) que en gens amb nombre variable de còpies no és evident el nombre de còpies que corresponen a cada al·lel en casos de mostres heterozigotes amb més de dues còpies; per exemple, una mostra que tingui 3 còpies de *CCL4L* i sigui heterozigota per a l'SNP rs4796195 (que presenta les variants al·lèliques L1 i L2) pot tenir: a) 2 còpies de l'al·lel L1 i 1 còpia de l'al·lel L2, o b) 1 còpia de l'al·lel L1 i 2 còpies de l'al·lel L2. Aquesta situació es pot complicar força més en casos de mostres heterozigotes amb 4, 5, 6 o més còpies de *CCL4L*. Per tant vam decidir que una de les fites importants d'aquest treball era establir un sistema de genotipificació que ens permetés determinar el nombre exacte de còpies de cada variant al·lèlica de *CCL4L*. En aquest cas un protocol de genotipificació que incloïa l'ús de sondes FRET (i no TaqMan) va fer-ho possible.

Tal i com es veurà en l'article, aquest treball ens ha permès tenir una visió clara i global de l'estructura poblacional pel què fa a la variabilitat de nombre de còpies de *CCL4L* i a la distribució de freqüències dels dos SNPs estudiats. El treball també posa en evidència la necessitat d'un protocol adequat per analitzar conjuntament CNVs i SNPs, especialment en estudis de susceptibilitat/resistència a malalties.



## Detailed Status Information

<b>Manuscript #</b>	2007-GGI-234
<b>Current Revision #</b>	0
<b>Submission Date</b>	27th Nov 07 10:46:33
<b>Current Stage</b>	Manuscript Submitted
<b>Title</b>	Population Structure in Copy Number Variation (CNV) and SNPs in the <i>CCL4L</i> Chemokine Gene
<b>Running Title</b>	CNV and SNPs in the <i>CCL4L</i> chemokine gene
<b>Manuscript Type</b>	Original Article
<b>Corresponding Author</b>	Manel Juan (Banc de Sang i Teixits (BST) / Universitat Autònoma de Barcelona (UAB))
<b>Contributing Authors</b>	Dr. Roger Colobran , David Comas , Rosa Faner , Edume Pedrosa , Roger Anglada , Ricardo Pujol , Jaume Betranpetit
<b>Financial Disclosure</b>	I have no relevant financial interests in this manuscript.
<b>Abstract</b>	The recent description of a large amount of copy number variation (CNV) in the human genome has extended the concept of genome diversity. In this study we integrate the analysis of CNV and SNPs in the human <i>CCL4L</i> chemokine gene. <i>CCL4L</i> is a non-allelic copy of <i>CCL4/MIP-1β</i> chemokine and displays a CNV that includes also the <i>CCL3L</i> gene, a non-allelic copy of <i>CCL3/MIP-1α</i> . This CNV and two functionally relevant <i>CCL4L</i> SNPs (rs4796195 and rs3744595) have been recently associated to HIV pathology in three independent studies. We have quantified the <i>CCL4L</i> copy number and genotyped both SNPs in samples from HGD-CEPH Diversity Panel. A strong correlation between <i>CCL4L</i> CNV and one of the SNPs analysed is found, whereas no significant linkage disequilibrium is found between the two SNPs despite their close distance (647 bp), suggesting a recent appearance of the second SNP when the diversity in the first one and CNV had already been generated. The present study points out that, in genes with CNV, it may be a key issue to combine the assessment of gene copy number with the genotyping of relevant SNPs to understand the phenotypic impact of genome variation in the immune response.
<b>Editor</b>	<a href="#">Grant Gallagher</a>
<b>Keywords</b>	Human, Chemokines, Genomics, Copy Number Variation

Stage	Start Date
Under Consideration	27th Nov 07 10:46:33
Author Approved Converted Files	27th Nov 07 10:46:33
Waiting for Author Approval of Converted Files	27th Nov 07 10:44:31
Submission	27th Nov 07 10:23:11
Manuscript Files Submitted	27th Nov 07 10:23:11
Preliminary Manuscript Data Submitted	27th Nov 07 10:17:00





## TITLE PAGE

### Population Structure in Copy Number Variation (CNV) and SNPs in the *CCL4L* Chemokine Gene<sup>1</sup>

Running title: CNV and SNPs in the *CCL4L* chemokine gene

Roger Colobran\*, David Comas<sup>†</sup>, Rosa Faner\*, Edurne Pedrosa\*, Roger Anglada<sup>‡</sup>, Ricardo Pujol-Borrell\*, Jaume Bertranpetit<sup>†</sup>, Manel Juan<sup>2\*</sup>.

\*Laboratori d'Immunobiologia per a la Recerca i Aplicacions Diagnòstiques (LIRAD), Banc de Sang i Teixits (BST), Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP), Badalona, Barcelona, Spain and Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain.

<sup>†</sup>Unitat de Biologia Evolutiva. Departament de Ciències Experimentals i de la Salut. Universitat Pompeu Fabra (UPF), Barcelona, Spain.

<sup>‡</sup>Servei de Genòmica. Departament de Ciències Experimentals i de la Salut. Universitat Pompeu Fabra (UPF), Barcelona, Spain.

Keywords: Human, Chemokines, Molecular Biology, AIDS.

## ABSTRACT

The recent description of a large amount of copy number variation (CNV) in the human genome has extended the concept of genome diversity. In this study we integrate the analysis of CNV and SNPs, combining the assessment of gene copy number with the genotyping of relevant SNPs in human *CCL4L* chemokine gene. *CCL4L* is a non-allelic copy of *CCL4/MIP-1 $\beta$*  chemokine and displays a CNV that includes also the *CCL3L* gene, a non-allelic copy of *CCL3/MIP-1 $\alpha$* . This CNV and two functionally relevant *CCL4L* SNPs (rs4796195 and rs3744595) have been recently associated to HIV pathology in three independent studies. We have quantified the *CCL4L* copy number and genotyped both SNPs in samples from HGDP-CEPH Diversity Panel. For the first time, we report worldwide population data combining both types of variation, CNV and SNPs, and our results show a clear population differentiation which is three times greater in the CNV than in the SNPs in *CCL4L* gene. A strong correlation between *CCL4L* CNV and one of the SNPs analysed is found, whereas no significant linkage disequilibrium is found between the two SNPs despite their close distance (647 bp), suggesting a recent appearance of the second SNP when the diversity in the first one and CNV had already been generated. The present study points out that, in genes with CNV, it may be a key issue to combine the assessment of gene copy number with the genotyping of relevant SNPs to understand the phenotypic impact of genome variation in the immune response.

## INTRODUCTION

Human diversity at the DNA level ranges from structural variation (such as duplications, deletions, insertions and inversions) to single nucleotide polymorphisms (SNPs). Until recently, SNPs have been the dominating type of variation explored in the human genome since it had been thought that SNPs account for much of the normal phenotypic diversity. However, the presence of copy number variation (CNV) in normal individuals has been recently reported as the major type of genetic diversity among humans (1, 2), accounting for 12% of the genome (3, 4). CNV involves DNA fragments about one kilobase or larger and, despite their heterogeneous distribution throughout the genome, there are no large stretches of the genome exempt from CNV (5). Nonetheless, fine sequence analyses among the members of a CNV have only been reported for very specific cases known before the CNV concept, such as colour-vision or immunologic functions (6, 7). Several examples of CNV affecting genes of the immune system have been previously described (e.g.  $\alpha/\beta$ -defensins and FcgrIII) (7-9). From the broad HLA polymorphisms to antigen receptor diversity by recombination, molecular differences form the basis for the efficiency of the immune response. In fact, the diversity of some gene families in the immune system has become a model of molecular diversity. One of these families is the chemokine superfamily, small structurally related cytokines that have evolved to form, in humans, a complex network of proteins able to carry out many diverse immune/non-immune and inflammatory/homeostatic functions (10, 11). Most, if not all, chemokines probably arose by gene duplication from a single ancestral gene and this fact is clearly reflected in the genomic organization of inflammatory chemokines: these genes are grouped in discrete chromosomal locations forming clusters and miniclusters that are considered single entities based on their overall function (12, 13).

*CCL3* (MIP-1 $\alpha$ ) and *CCL4* (MIP-1 $\beta$ ) constitute, together with *CCL18* (PARC), an interesting miniclust model of chemokine evolution located on chromosome 17q12. *CCL3* and *CCL4* genes were formed by duplication of an ancestral gene and *CCL18* is an unusual large chemokine gene generated by fusion of two *CCL3*-like genes (14, 15). Additionally, *CCL3* and *CCL4* both have a second non-allelic copy, *CCL3L* (LD78 $\beta$ ) and *CCL4L* (LAG-1), that code for two highly similar proteins (>90% identity between *CCL3* and *CCL3L* proteins, and >95% between *CCL4* and *CCL4L* proteins) (16, 17). *CCL3L* and *CCL4L* exhibit copy number variation (CNV), meaning that these genes are present in variable copy number in the human genome (18). Regarding the CNV found in the *CCL3L* and *CCL4L* genes, it seems to have been generated by duplications of a 120 kb stretch in this region of chromosome 17 (including *CCL3L*, *CCL4L* and other non-related loci) (19, 20). However, two key points should be emphasized: i) there are some individuals that lack the *CCL3L-CCL4L* tandem, and ii) there are haplotypes with different gene copy number (meaning that the gene copy number of both genes can be different in a single chromosome) (18). Evidence of a functional significance for the *CCL3L* CNV has been described by Gonzalez et al. demonstrating their influence on HIV susceptibility (21). According to Gonzalez et al., the strength of the immune response based on *CCL3L* activity depends on the relative gene copy number of the individual in relation to the mean copy number of the ethnic group to which the individual belongs. This fact has two relevant implications: i) there is a complex ethnic-specific adaptation of different factors in the immune response, and ii) the knowledge of the population-specific variation and the worldwide distribution of copy number should be approached in order to unravel individual susceptibility and response to pathogens.

In the case of *CCL4L*, and in addition to the CNV, a recently described SNP (rs4796195) in the acceptor splice site of intron-2 generates a new complex splicing pattern (19). This polymorphism creates two allelic variants: *CCL4L\*1*, the original variant (that will

be called L1), and the new described variant, *CCL4L*\*2 (or L2 allele) which has been associated with HIV susceptibility in a case-control study (19). Additionally, another SNP in the exon-2 of *CCL4L* gene (rs3744595) leads to a non-conservative amino acid change (Arg22His), likely to have functional implications. These two alleles will be called R for Arginine and H for Histidine. Initial evidence suggest that this polymorphism can influence the survival of HIV+ patients (22). These previous studies do not take into account the *CCL4L* CNV and, in fact, despite the growing interest in the identification of these CNVs and their phenotype implications to the individual structural variation found in humans (3, 23), no data combining both types of polymorphisms is available, and the extent of independence between both types of variation (CNVs and SNPs) is still unknown.

The main aim of this study is to combine and correlate the information provided by the CNV and the SNPs in the *CCL4L* gene, analyzing their population structure and variation, a key factor in understanding individual immune response. We have analyzed the worldwide distribution of *CCL4L* copy number and genotyped two relevant SNPs determining the distribution of alleles in heterozygous multicopy samples, showing their independent structure and complexity, that cannot be ignored when analysing the impact of chemokines in the response to pathogens.

## MATERIALS AND METHODS

### Samples

A total of 1,064 individuals from 52 worldwide populations included in the Human Genome Diversity Panel (HGDP) (24) were analysed for the *CCL4L* copy number determination and genotyped for the SNPs selected. From the original panel, several duplicated samples were used as internal controls and some atypical individuals were ignored. The panel used is the H1048 according to Rosenberg et al. which includes 1,048 individuals (25). For some of the analyses, individuals were pooled into seven broad continental regions: sub Saharan Africa (SSAFR), Middle East/North Africa (MENA), Europe (EUR), Central/South Asia (CSASIA), East Asia (EASIA), Oceania (OCE) and America (AME).

### *CCL4L* copy number determination

*CCL4L* copy number determination was performed by using real-time PCR (rt-PCR) in an ABI PRISM® Sequence Detection System 7900 Instrument (Applied Biosystems). Using the Custom Taqman® Gene Expression Assay Service (Applied Biosystems), a mix of unlabeled PCR primers and a TaqMan® MGB probe, FAM™ dye-labeled was designed to amplify specifically *CCL4L* and clearly discriminate *CCL4L* from *CCL4*. Primers and probe sequences are as follows: sense primer (located at intron 2) 5'-CATGGTCAGGCAGAGGAAGATG- 3'; antisense primer (located at exon 3) 5'-GCTTGCCTCTTTTGGTTTGGAAAT- 3'; probe (located at intron 2, in a region that is deleted in *CCL4*) 5'-TACCACAGGCAAGGGAT- 3' (FAM labeled). As a control, TaqMan® *RNaseP* Control (Applied Biosystems) was used following supplier's recommendations. The amplification protocol for this reaction was: 10 minutes of initial setup at 95°C, followed by 40 amplification cycles (15 seconds of denaturation at 95°C and

60 seconds of annealing/extension at 60°C). For each sample, amplification of the target gene (*CCL4L*) and a control gene (*RNaseP*, that has 2 copies per diploid human genome) was performed in the same well (multiplex format) and by duplicate. To generate standard curves we used six serial 1:2 dilutions (146.4 – 4.57 ng) of genomic DNA from CEM cells (known to have two copies of *CCL4L* per diploid genome as proved by Southern blot densitometry) (18). Each standard curve dilution was run in triplicate in each PCR plate (384 wells) for *CCL4L* and *RNaseP* using the primers and probes described above (multiplex). The square of the Pearson correlation coefficient ( $R^2$ ) for a standard curve was  $>0.97$  (in case of lower value, samples would have been repeated, although in our hands this situation was not encountered). Threshold cycle (CT), the cycle number at which the fluorescence reaches a fixed threshold, of each sample was determined and converted into template quantity using the standard curves. Copy number was established as the ratio of the template quantity for *CCL4L* to the template quantity for *RNaseP*, multiplied by two. Average copy number of duplicates was calculated based on the ratio between sample and *RNaseP* single locus control. We also calculated the standard deviation (SD) and which percentage represented the SD with reference to the average, accepting only values  $<10\%$  (in case of greater differences, quantification of samples was repeated). Once all these steps were successfully done, we assigned a real number of copies to each sample. We established a difference of  $\pm 0.25$  as the threshold to consider a result as a real number of copies (for example, a final number of 2.20 or 1.80 copies is considered as 2 real copies). This value, according to the distribution found, is very conservative, with a negligible number of cases out of the established range. When the final number exceeded the established threshold, the quantification of the sample was repeated.

### ***CCL4L* SNPs genotyping**



*CCL4L* genotyping for rs4796195 and rs3744595 SNPs was performed using rt-PCR with fluorescence resonance energy transfer (FRET) probes in a LightCycler Instrument (Roche, Mannheim, Germany). For each SNP we used a pair of specific primers for the *CCL4L* region where the SNP is located and a pair of FRET probes. Primers and probes sequences for rs4796195 SNP are as follows: sense primer (located at intron 2) 5'-GCAGAGGAAGATGCCTACCAC- 3'; antisense primer (located at exon 3) 5'-CTGAGTATGGAGGAGATGCGG- 3'; anchor fluorescein-labeled probe (located at exon 3) 5'-AGCGCAGACTTGCTTGCCTCTTTT- 3'; sensor LC-Red705-labeled probe (located at intron 2/exon 3 junction) 5'-TTTGGAAATCCGTAGAACAAGG- 3'. Primers and probes sequences for rs3744595 SNP are as follows: sense primer (located at intron 1) 5'-GGAATGGATACAAGGGACCA- 3'; antisense primer (located at intron 2) 5'-CAGCCAGGGGTTGATACTC- 3'; anchor fluorescein-labeled probe (located at exon 2) 5'-CAGAGGCTGCTGGTCTCATAGTAATCT- 3'; sensor LC-Red610-labeled probe (located at exon 2) 5'-CCACAAAGTTGCGAGGAA- 3'. PCR reaction was performed in a final volume of 10 ml; the reaction mixture contained 1 ml of LightCycler® FastStart DNA Master HybProbe, 1mM primer sense, 0.2 mM primer antisense, 0.2 mM of each hybridization probe (anchor and sensor), 2 mM of MgCl<sub>2</sub> and 2 ml of template DNA (the amount of DNA added to each PCR reaction was between 10-20 ng). The amplification protocol for these reactions was: 10 minutes of initial denaturation and activation of the FastStart enzyme at 95°C, followed by 45 amplification cycles (5 seconds of denaturation at 95°C, 6 seconds of annealing at 55°C and 10 seconds of extension at 72°C). Melting curves were generated following these steps: 95°C for 0 seconds, 65°C for 15 seconds, 42°C for 270 seconds (all at a ramp rate of 20°C/second) and 80°C for 0 seconds (ramp rate of 0.05°C/second; acquisition mode: continuous), followed by a cooling step of 30 seconds at 40°C.

## **Numerical allelic discrimination in heterozygous individuals for rs4796195 and rs3744595 SNPs**

When two allelic variants are detected during the SNP genotyping process, it is not straightforward, except in cases of two *CCL4L* copies, to determine how many *CCL4L* copies present each allelic variant. To assess the exact number of *CCL4L* copies to each allelic variant in heterozygous samples with more than two *CCL4L* copies, we used the melting curves produced by FRET probes. Using the chart that plots the first negative derivative of the sample fluorescent curves, the melting temperature of each sample appears as a peak. Displaying the melting temperatures as peaks makes it easier to distinguish each sample's characteristic melting profile and to discern differences between samples. Our genotyping analysis result data included each sample's melting temperature and also the size of the area under each sample's melting peak. For each SNP, each allelic variant showed a characteristic  $T_m$ : for SNP rs4796195, the L1 and L2 variants show a  $T_m$  of  $53\pm 0.5^\circ\text{C}$  and  $62\pm 0.5^\circ\text{C}$  respectively, whereas the SNP rs3744595, the R (Arginine) and H (Histidine) variants showed a  $T_m$  of  $55.5\pm 0.5^\circ\text{C}$  and  $48\pm 0.5^\circ\text{C}$  respectively. In case of heterozygous samples, the ratio between the areas' peaks allowed us to ascribe the exact number of *CCL4L* copies to each allelic variant.

### **Statistical analysis**

Data sets were analyzed using Sigmastats software. When necessary, the results were expressed as the mean value  $\pm$  SD. A Pearson correlation test was applied to the data sets to determine statistically significant correlation differences.

In order to assess the population structure,  $F_{ST}$  values based on the individual CNV were calculated with the ARLEQUIN package ver 2.000 (26). In order to apportion the fraction of the genetic variance due to differences between and within continental groups,

several analyses of the molecular variance (AMOVA) (27), were performed also using Arlequin. AMOVA was performed both for the whole population set and for each of the seven continental groups previously defined.

Linkage disequilibrium (LD) between rs4796195 and rs3744595 SNPs was measured with  $D'$  and  $r^2$  parameters with the Haploview software. Since CNV did not allow us to phase unambiguously the alleles of both SNPs, LD was calculated taking into account the 149 individuals who exhibit a single copy of *CCL4L*.

## RESULTS

### Worldwide *CCL4L* copy number distribution and population structure

*CCL4L* copy number has been determined in the 1,048 individuals included in the HGDP panel. Similar results were obtained analyzing smaller sets (H971 and H952) of individuals (data not shown) (25). Results for each individual population are shown in Figure 1. The similarity of the population CNV distribution between *CCL4L* and *CCL3L* performed in the same set of individuals is noteworthy (21). To give a more general view of the worldwide distribution of *CCL4L* copy number, individuals were pooled into seven broad continental regions and the mean of *CCL4L* copy number was calculated. Sub Saharan African populations display the highest number of *CCL4L* copies (mean =  $4.32 \pm 0.63$ , including individuals with 8, 9 or 10 copies), whereas Europe present the lowest copy number ( $1.89 \pm 0.40$ ), including a 25% of individuals with only one *CCL4L* copy. The number of individuals without *CCL4L* gene was always below 5% in all continental regions (Sub-Saharan Africa 0%, America 0%, East Asia 1.82%, Europe 2.50%, Central/South Asia 2.62%, Oceania 2.63% and Middle East/North Africa 3.41%).

In order to test the population differentiation concerning the *CCL4L* copy number, the  $F_{ST}$  statistic was calculated for the whole set of populations. A significant 0.12 ( $p < 0.001$ )  $F_{ST}$  value was found, a value very similar to the average for 67 autosomal CNVs (0.11) previously described in the four HapMap populations (4). When an analysis of the molecular variance (AMOVA) was performed grouping the populations within the seven broad continental regions, 9.26% ( $p < 0.001$ ) of the variation found in the copy number was the result of differences between continental groups, and only 3.43% ( $p < 0.001$ ) were due to differences within continental groups. This general heterogeneity found within continental groups was significant in Asian (East Asia = 4.63%,  $p < 0.001$ ; Central/South Asia = 4.32%,

p=0.004) and American populations (8.75%, p=0.002), whereas no significant differences were found between populations within the rest of continental groups including Africa, which is noteworthy as most other genetic markers show strong heterogeneity within the African continent.

### **Population distribution of SNPs allelic and genotypic frequencies**

One of the technical challenges in analysing SNPs within genes with CNV is the need to establish the exact number of copies for each individual, then perform the SNP genotyping and finally determine the numeric distribution of the allelic variants. The melting curves characteristic for each of the two SNPs genotyped allowed us to quantify the number of copies of each allelic variant for each individual. Figure 2 (A and B) shows the allelic discrimination in copy number for L1 and L2 variants. Analysing the melting curve profiles in different individuals with the same ratio of allelic copies we provided evidence of the procedure reliability (figure 2C). The same rationale was used to distinguish R and H variants (data not shown).

L1/L2 and R/H allelic distributions are shown in Figure 3. L2 frequencies (Figure 3A) range from 13% in Oceania to 43% in the Americas, being polymorphic in all the populations analysed. On the other hand, the H frequency is very low in most of the populations analysed, the R allele being fixed in some of the samples, and almost not present in sub Saharan Africa and America. Concerning the population structure of these SNPs in the dataset, an  $F_{ST}$  value of 0.04 was found for both polymorphisms (p<0.001), a smaller value than the one found for the *CCL4L* CNV (0.12). When the populations were grouped according to continental areas, 3.82% (p<0.001) and 3.39% (p<0.001) of the genetic variation for the L1/L2 and the R/H alleles respectively, was due to differences between continental regions. Nevertheless, the genetic heterogeneity within continents was very low and only significant in sub Saharan

Africa (3.51%,  $p=0.002$ ) and Oceania (8.29%,  $p=0.043$ ) for the L1/L2 polymorphism, and in America (1.94%,  $p=0.029$ ) for the R/H polymorphism.

The genotypic frequencies were also calculated considering that, in the case of genes with CNV, what is usually called “heterozygous” are those that present the two allelic variants in their genome regardless of their copy number (Figure 4). A relevant aspect is that in genes with CNV the relationship between allelic and genotypic frequencies do not necessarily maintain the Hardy-Weinberg equilibrium (HWE) due to the unbalanced distribution of SNP variants among different copies of the gene in the same chromosome. For L1/L2 genotypes, the detailed frequencies for each genotype in individual populations and continental regions are shown in figure 4A. Interestingly, whereas the maximum heterozygosity frequency expected following the HWE is 50%, there are three continental regions that show a frequency over 50%: East Asia (66%), sub Saharan Africa (85%) and America (86%), including specific populations of sub Saharan Africa and America with heterozygosity values over 90%. On the other hand, R/H genotypes for continental regions follow the HWE except for the Oceanic populations (figure 4B).

### **Correlation between CNV and SNPs in the *CCL4L* gene**

To better define the relationship between CNV and SNPs in the *CCL4L* gene, we analyzed the correlation between the *CCL4L* copy number and the frequencies of the two analyzed SNPs using two different approaches: a) *CCL4L* copy number versus minor allele frequency and b) *CCL4L* copy number versus heterozygous frequency (Figure 5). For the L1/L2 copies, there is a strong correlation between *CCL4L* copy number and allelic and heterozygous frequency ( $R=0.731$ ,  $p<0.0001$  and  $R=0.890$ ,  $p<0.0001$  respectively), pointing to a common evolutionary history of these polymorphisms. For the R/H copies, a small negative correlation between *CCL4L* copy number and allelic frequency ( $R=-0.491$ ,

$p=0.0002$ ) and a lack of correlation between CNV and heterozygous frequency ( $R=-0.186$ ,  $p=0.190$ ) were found, suggesting that these two kinds of polymorphisms arose independently.

The opposite results of the two analyzed SNPs regarding to *CCL4L* CNV is in agreement with the lack of linkage disequilibrium between them ( $r^2=0.0185$ ;  $p=0.097$ ), which has been calculated using data of individuals with one *CCL4L* copy ( $n=149$ ). The haplotypes found in these set of individuals consisted of 111 L1/R copies, 26 L2/R copies and 12 L1/H copies. No L2/H copies were found in the set analysed, this leads to a linkage disequilibrium measure of  $D'=1$ . The lack of linkage disequilibrium between both SNPs, which are at 647 bp distance, and the lack of correlation between CNV and L/H copies are noteworthy.

## DISCUSSION

The study of the functional implications of genetic diversity has generated a large amount of data involving SNPs in many pathologies as well as in non-pathological processes (28-30). Recently, the widespread presence of CNVs in the genomes of healthy individuals with no obvious genetic disorders has been described in a few populations (1, 2), but relatively few data has been reported regarding CNVs and disease resistance or susceptibility (9, 21). Regarding this point, the main aim of our work was to link the analysis of SNPs and CNVs within a worldwide framework to achieve a more powerful tool in the future determination of genetic basis of susceptibility or resistance to disease. It seems clear that genes with CNV require a non-conventional method to study their SNPs, and this is the reason why we established a clear and precise approach to solve this problem. We focused on the *CCL4L* gene, which provides one of the best examples of CNV in the human immune system (*CCL4L* copies in our population set range from 0 to 10) combined with the presence of interesting SNPs (rs4796195 and rs3744595 both related with HIV pathology) and the study was carried out in a large set of populations to provide a global idea of the contribution of each element and to detect the possible interactions between them. The knowledge of the intra- and inter-population variation seems to be the key parameter in understanding the individual chemokine immune response, beyond the knowledge of the individual genetic composition (21).

The CNV analyzed in this study belongs to the multi-allelic CNVs (see CNV classification in (4)) and extends across a 120 kb stretch in the q12 region of chromosome 17 that includes *CCL4L* and *CCL3L* genes. The high level of similarity between them, common gene strand orientations, comparable intergenic distances, and amino acid similarities strongly argue for non-allelic recombination as a mechanism responsible for the origin of one



of these gene pairs (20). Complex patterns of copy number variation have been described at sites of segmental duplications with several genes involved in the repeated segment (31, 32). To date, *CCL4L* and *CCL3L* are the only chemokine genes that are present in the human population at variable number of copies. Interestingly, while there are often equivalent numbers of *CCL3L* and *CCL4L* genes, there are some exceptions suggesting that these genes are not simply coduplicated. Recently, Gonzalez et al. (21), quantified the *CCL3L* copy number in the same individual cohort (HGDP) and, although the *CCL3L* raw data is not available for comparison, *CCL4L* results are highly concordant, although a slight lower copy average of *CCL4L* gene is found compared to *CCL3L* in all populations and continental regions. This high correlation between *CCL4L* and *CCL3L* copies strongly suggests the involvement of both genes in the HIV susceptibility. The lower average of *CCL4L* copies found in the present analysis is in agreement with the results of Towson et al. (18), who quantified the *CCL3L* and *CCL4L* gene copies in 100 individuals of European ancestry, showing a 60% of samples with equal number of *CCL3L* and *CCL4L* genes, 33% having more *CCL3L* copies compared with *CCL4L* (generally one or two copies more) and only a 7% of samples having more *CCL4L* copies. All these results support the presence of a single unit containing only one of these genes and being also susceptible to duplication. Thus, the consideration of both genes as a single repeat unit is an oversimplification, as it is also the consideration of the raw number of repeats in the *CCL4L* region. This additional complexity found in the *CCL3L-CCL4L* CNV implies that it is necessary to check if all the genes spread through a region with CNV are the result of a single duplication/deletion process.

The *CCL4L* CNV described shows a clear population differentiation, similar to that previously found on 67 autosomal CNVs with the HapMap samples. Besides this continental differentiation in CNV, Asian and Amerindian populations exhibit a genetic heterogeneity in copy number whereas the rest of the continental groups are genetically homogeneous,

including Africans. It is noteworthy that the *CCLAL* genetic differentiation between populations is three times larger in CNV ( $F_{ST} = 0.12$ ) than in the SNPs analysed ( $F_{ST} = 0.04$ ). Thus, the population differentiation in relation to SNPs frequencies within the *CCLAL* copies is more homogeneous between and within continental regions than CNV. This result suggests that the analysis of CNV might be more informative in unravelling population genetic relationships than SNP variation but it does not allow a direct functional correlation. Moreover, it suggests complex dynamics in the region as it cannot be explained as a variation in the number of copies of pre-existing different units defined by their nucleotide composition.

With regard to the SNP analysis, despite the relatively short distance between them (647 bp), they are not in linkage disequilibrium (LD) and their worldwide distribution is very different. The low frequency of the H allele worldwide, the absence of L2/H haplotypes, and the lack of correlation between R/H polymorphism and CNV suggests that the H allele appeared recently in a L1 copy background. Our results are compatible with the appearance of the allele H once the L1/L2 diversity was already established. This would imply that the H variant is completely associated with the L1 copy where it was generated. This fact would explain why the H variant does not correlate neither with the CNV nor with L1/L2 variation despite the short distance between both SNPs. It is also interesting to note the near absence of the H variant in sub Saharan African samples, which reinforces the idea of the recent appearance of the H allele in human populations.

Association studies have focused primarily on SNP genotyping. However, it is important to determine whether CNVs are in linkage disequilibrium with common SNPs, and thus if CNV could be assessed indirectly in SNP-based studies. Regarding this question, although it has been shown that small indels and surrounding SNPs are in linkage disequilibrium in the human genome (33), Redon et al. found that diploid copy number of

multi-allelic CNVs is poorly predicted by neighbouring SNPs (4). Interestingly, among the populations included in this study the L1/L2 variants (located inside the CNV) have a high correlation with the *CCL4L* copy number, and, therefore, these variants could be used as a proxy of the individual *CCL4L* copies. The possibility to determine the copy number of a gene or group of genes through the genotyping of a SNP is an attractive approach since, to date, SNP genotyping is an easier and simpler methodology than copy number determination.

As previously mentioned, the three *CCL4L* genetic variation elements (the *CCL4L* CNV and both SNPs analysed) are involved in HIV pathology. However, each of them was analysed independently in previous studies and it would be relevant to determine the specific contribution of each one in susceptibility/resistance to HIV. Data of our study revealed that sub Saharan African and Amerindian populations display the highest number of *CCL4L* (and *CCL3L*) copies (described as a HIV resistance factor) (21) but, at the same time, showed also the highest frequency of the L2 allele (described as a HIV susceptibility factor) (19). Conversely, the European populations display the lowest number of *CCL4L-CCL3L* copies and also a low frequency of L2 allele. Additionally, the lack of correlation of the R/H variants with *CCL4L* copy number increase the complexity in the interpretation of the role of these three elements. The balance between presenting a high or low number of *CCL4L-CCL3L* copies (compared to the population distribution) and the L2 allele frequency, together with all the other HIV susceptibility/resistance described factors (including the rs3744595 SNP role in survival of HIV+ patients), could determine the final contribution of each element to this pathology.

In genes with CNV it is important to combine the assessment of gene copy number with the genotyping of relevant SNPs and it should now become standard in the design of all studies of the genetic basis of phenotypic variation, in particular in studies of resistance/susceptibility to disease. Moreover, the population variation is, in some cases, a key factor to understand

the individual susceptibility. We envisage that this integrated analysis will result in a better comprehension of these two most important genetic variability generating mechanisms.

## **ACKNOWLEDGMENTS**

We thank Maria del Pilar Armengol and Michelle Gardner for critical reading of the manuscript. We thank Howard Cann (CEPH, Paris) for providing the HGDP-CEPH panel. We also thank Mònica Vallés (UPF) and Stéphanie Plaza (Servei de Genòmica, UPF) for technical assistance.

## **DISCLOSURES**

The authors declare no conflict of interest or financial interests.

## REFERENCES

1. Iafrate, A. J., L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, S. W. Scherer, and C. Lee. 2004. Detection of large-scale variation in the human genome. *Nat Genet* 36:949.
2. Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, and M. Wigler. 2004. Large-scale copy number polymorphism in the human genome. *Science* 305:525.
3. Freeman, J. L., G. H. Perry, L. Feuk, R. Redon, S. A. McCarroll, D. M. Altshuler, H. Aburatani, K. W. Jones, C. Tyler-Smith, M. E. Hurles, N. P. Carter, S. W. Scherer, and C. Lee. 2006. Copy number variation: new insights in genome diversity. *Genome Res* 16:949.
4. Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, A. J. Iafrate, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles. 2006. Global variation in copy number in the human genome. *Nature* 444:444.
5. Wong, K. K., R. J. deLeeuw, N. S. Dosanjh, L. R. Kimm, Z. Cheng, D. E. Horsman, C. MacAulay, R. T. Ng, C. J. Brown, E. E. Eichler, and W. L. Lam. 2007. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* 80:91.

6. Vollrath, D., J. Nathans, and R. W. Davis. 1988. Tandem array of human visual pigment genes at Xq28. *Science* 240:1669.
7. Fellermann, K., D. E. Stange, E. Schaeffeler, H. Schmalzl, J. Wehkamp, C. L. Bevins, W. Reinisch, A. Teml, M. Schwab, P. Lichter, B. Radlwimmer, and E. F. Stange. 2006. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 79:439.
8. Linzmeier, R. M., and T. Ganz. 2005. Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics* 86:423.
9. Aitman, T. J., R. Dong, T. J. Vyse, P. J. Norsworthy, M. D. Johnson, J. Smith, J. Mangion, C. Robertson-Lowe, A. J. Marshall, E. Petretto, M. D. Hodges, G. Bhangal, S. G. Patel, K. Sheehan-Rooney, M. Duda, P. R. Cook, D. J. Evans, J. Domin, J. Flint, J. J. Boyle, C. D. Pusey, and H. T. Cook. 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 439:851.
10. Baggiolini, M. 1998. Chemokines and leukocyte traffic. *Nature* 392:565.
11. Luster, A. D. 1998. Chemokines--chemotactic cytokines that mediate inflammation. *N Engl J Med* 338:436.
12. Zlotnik, A., and O. Yoshie. 2000. Chemokines: a new classification system and their role in immunity. *Immunity* 12:121.
13. Colobran, R., R. Pujol-Borrell, M. P. Armengol, and M. Juan. 2007. The chemokine network. I. How the genomic organization of chemokines contains clues for deciphering their functional complexity. *Clin Exp Immunol* 148:208.
14. Menten, P., A. Wuyts, and J. Van Damme. 2002. Macrophage inflammatory protein-1. *Cytokine Growth Factor Rev* 13:455.



15. Tasaki, Y., S. Fukuda, M. Iio, R. Miura, T. Imai, S. Sugano, O. Yoshie, A. L. Hughes, and H. Nomiyama. 1999. Chemokine PARC gene (SCYA18) generated by fusion of two MIP-1alpha/LD78alpha-like genes. *Genomics* 55:353.
16. Menten, P., S. Struyf, E. Schutyser, A. Wuyts, E. De Clercq, D. Schols, P. Proost, and J. Van Damme. 1999. The LD78beta isoform of MIP-1alpha is the most potent CCR5 agonist and HIV-1-inhibiting chemokine. *J Clin Invest* 104:R1.
17. Modi, W. S., J. Bergeron, and M. Sanford. 2001. The human MIP-1beta chemokine is encoded by two paralogous genes, ACT-2 and LAG-1. *Immunogenetics* 53:543.
18. Townson, J. R., L. F. Barcellos, and R. J. Nibbs. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol* 32:3016.
19. Colobran, R., P. Adreani, Y. Ashhab, A. Llano, J. A. Este, O. Dominguez, R. Pujol-Borrell, and M. Juan. 2005. Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. *J Immunol* 174:5655.
20. Modi, W. S. 2004. CCL3L1 and CCL4L1 chemokine genes are located in a segmental duplication at chromosome 17q12. *Genomics* 83:735.
21. Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R. J. Nibbs, B. I. Freedman, M. P. Quinones, M. J. Bamshad, K. K. Murthy, B. H. Rovin, W. Bradley, R. A. Clark, S. A. Anderson, J. O'Connell R, B. K. Agan, S. S. Ahuja, R. Bologna, L. Sen, M. J. Dolan, and S. K. Ahuja. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307:1434.
22. Capoulade-Metay, C., L. Meyer, T. Tran, A. Persoz, A. Bourdais, Y. Dudoit, J. F. Delfraissy, P. Debre, and I. Theodorou. 2005. Influence of the R22H variant of macrophage inflammatory protein 1beta/Lag-1 in HIV-1 survival. *Aids* 19:831.
23. Sharp, A. J., Z. Cheng, and E. E. Eichler. 2006. Structural Variation of the Human Genome. *Annu Rev Genomics Hum Genet.*

24. Cann, H. M., C. de Toma, L. Cazes, M. F. Legrand, V. Morel, L. Piouffre, J. Bodmer, W. F. Bodmer, B. Bonne-Tamir, A. Cambon-Thomsen, Z. Chen, J. Chu, C. Carcassi, L. Contu, R. Du, L. Excoffier, G. B. Ferrara, J. S. Friedlaender, H. Groot, D. Gurwitz, T. Jenkins, R. J. Herrera, X. Huang, J. Kidd, K. K. Kidd, A. Langaney, A. A. Lin, S. Q. Mehdi, P. Parham, A. Piazza, M. P. Pistillo, Y. Qian, Q. Shu, J. Xu, S. Zhu, J. L. Weber, H. T. Greely, M. W. Feldman, G. Thomas, J. Dausset, and L. L. Cavalli-Sforza. 2002. A human genome diversity cell line panel. *Science* 296:261.
25. Rosenberg, N. A. 2006. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet* 70:841.
26. Schneider S, K. J.-M., Roessli D, Excoffier L. 1997. Arlequin version 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Genetics and Biometry Laboratory, University of Geneva, Geneva.
27. Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479.
28. Engle, L. J., C. L. Simpson, and J. E. Landers. 2006. Using high-throughput SNP technologies to study cancer. *Oncogene* 25:1594.
29. Fan, J. B., M. S. Chee, and K. L. Gunderson. 2006. Highly parallel genomic assays. *Nat Rev Genet* 7:632.
30. Eichler, E. E., D. A. Nickerson, D. Altshuler, A. M. Bowcock, L. D. Brooks, N. P. Carter, D. M. Church, A. Felsenfeld, M. Guyer, C. Lee, J. R. Lupski, J. C. Mullikin, J. K. Pritchard, J. Sebat, S. T. Sherry, D. Smith, D. Valle, and R. H. Waterston. 2007. Completing the map of human genetic variation. *Nature* 447:161.

31. Goidts, V., D. N. Cooper, L. Armengol, W. Schempp, J. Conroy, X. Estivill, N. Nowak, H. Hameister, and H. Kehrer-Sawatzki. 2006. Complex patterns of copy number variation at sites of segmental duplications: an important category of structural variation in the human genome. *Hum Genet* 120:270.
32. Sharp, A. J., D. P. Locke, S. D. McGrath, Z. Cheng, J. A. Bailey, R. U. Vallente, L. M. Pertz, R. A. Clark, S. Schwartz, R. Segraves, V. V. Oseroff, D. G. Albertson, D. Pinkel, and E. E. Eichler. 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 77:78.
33. Hinds, D. A., A. P. Kloek, M. Jen, X. Chen, and K. A. Frazer. 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet* 38:82.

## FOOTNOTES

<sup>1</sup>This work was supported by grants from Fondo de Investigaciones Sanitarias (Project 02/0104), Fundación para la Investigación y la Prevención del SIDA en España (FIPSE, Project 36487/05), and Direcció General de Recerca, Generalitat de Catalunya (SGR2005/00608).

<sup>2</sup>Address correspondence and reprint requests to Dr. Manel Juan. Postal address: Hospital Germans Trias i Pujol, Edifici de Recerca, Carretera de Can Ruti, camí de les escoles, s/n, 08916 Badalona, Spain. Telephone number: +34.93.497.86.66. FAX number: +34.93.497.86.68. E-mail address: [Manel.Juan@comb.es](mailto:Manel.Juan@comb.es)

## **ABBREVIATIONS**

Abbreviations used in this paper: CNV, copy number variation; SNP, single nucleotide polymorphism.

## FIGURE LEGENDS

### FIGURE 1. *CCL4L* copy number in worldwide populations

Results are shown as mean (column) and SD (bar) for each individual population.

Populations have been grouped into seven broad continental regions and the *CCL4L* mean copy number and the standard deviation is shown in brackets.

### FIGURE 2. L1 and L2 allelic discrimination in heterozygous individuals

(A) Melting curve profiles for rs4796195 genotyping. The first negative derivative of the sample fluorescent curves is represented. Each allelic variant (L1 and L2) shows a peak with a characteristic melting temperature. The size of each peak is related to the number of copies of *CCL4L* that display each allelic variant. Samples with the same copy number of each allelic variant (N:N) show equivalent peaks. Additional examples of different observed combinations are shown. (B) Numerical calculation to ascribe the exact number of *CCL4L* copies to each allelic variant using the ratio between the areas' peaks. Samples with the same copy number of each allelic variant (N:N) have a ratio of 1. Additional examples of different combinations are shown. (C) Reliability and reproducibility of allelic copy number determination for L1 and L2 allelic variants. Different examples of same ratios from different heterozygous individuals are shown as overlapped curves. Each curve represents an individual and the total number of *CCL4L* copies of each individual is given as the sum of L1 plus L2 copies.

### FIGURE 3. L1/L2 and R/H allelic frequencies in worldwide populations

Results for each individual population (bar graphics) and continental region (circular graphics) are shown.

**FIGURE 4. L1/L2 and R/H genotypic frequencies in worldwide populations**

Results for each individual population (bar graphics) and continental region (circular graphics) are shown. Individuals lacking the *CCL4L* gene are included. “(n)” means the presence of each allele in a variable copy number.

**FIGURE 5. Correlation between *CCL4L* copies and L1/L2 and R/H polymorphisms**

Each dot represents an individual population. Two types of correlations are shown: A) and B) *CCL4L* copies versus minor allele frequencies for L1/L2 and R/H respectively; and C) and D) *CCL4L* copies versus heterozygous frequencies for L1/L2 and R/H respectively. Two populations (Namibia and New Guinea) have been excluded of this analysis due to their outlier positions.

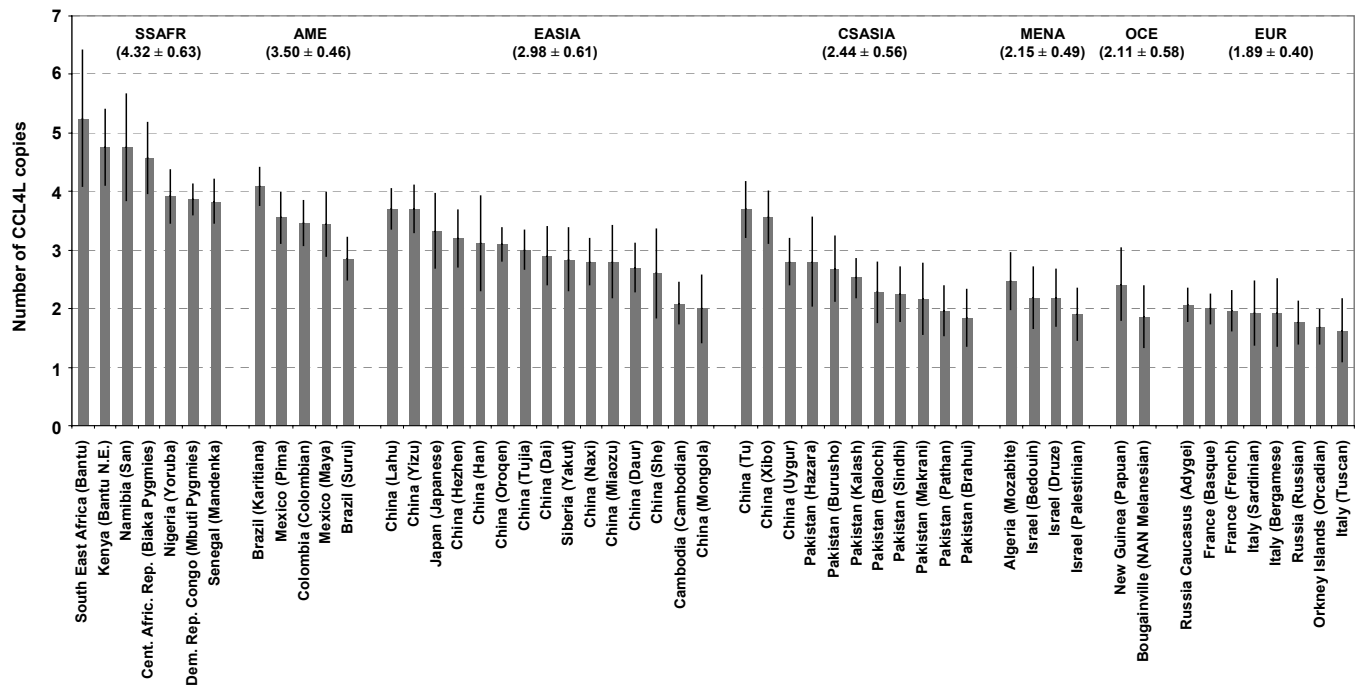


Figure 1



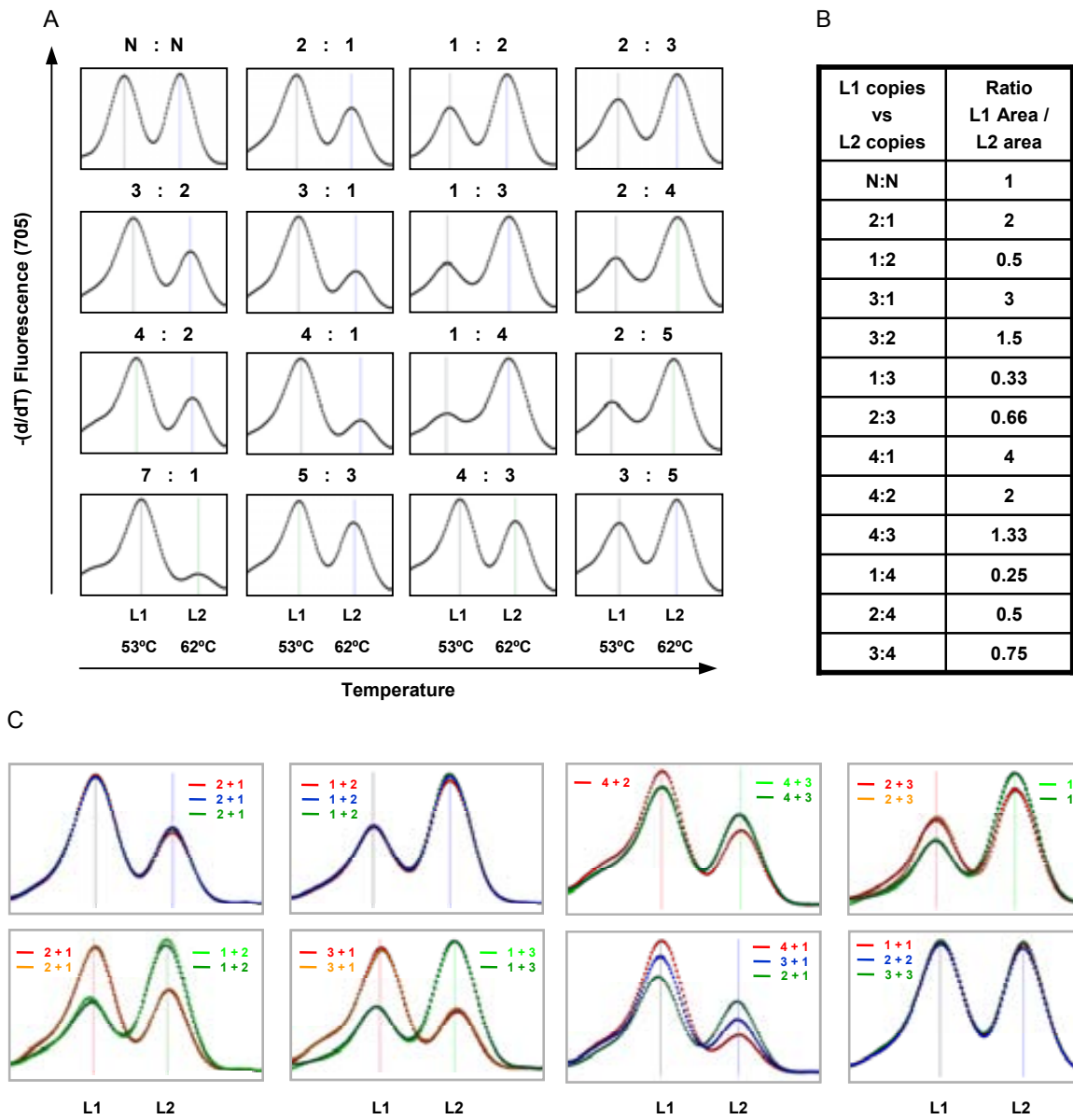
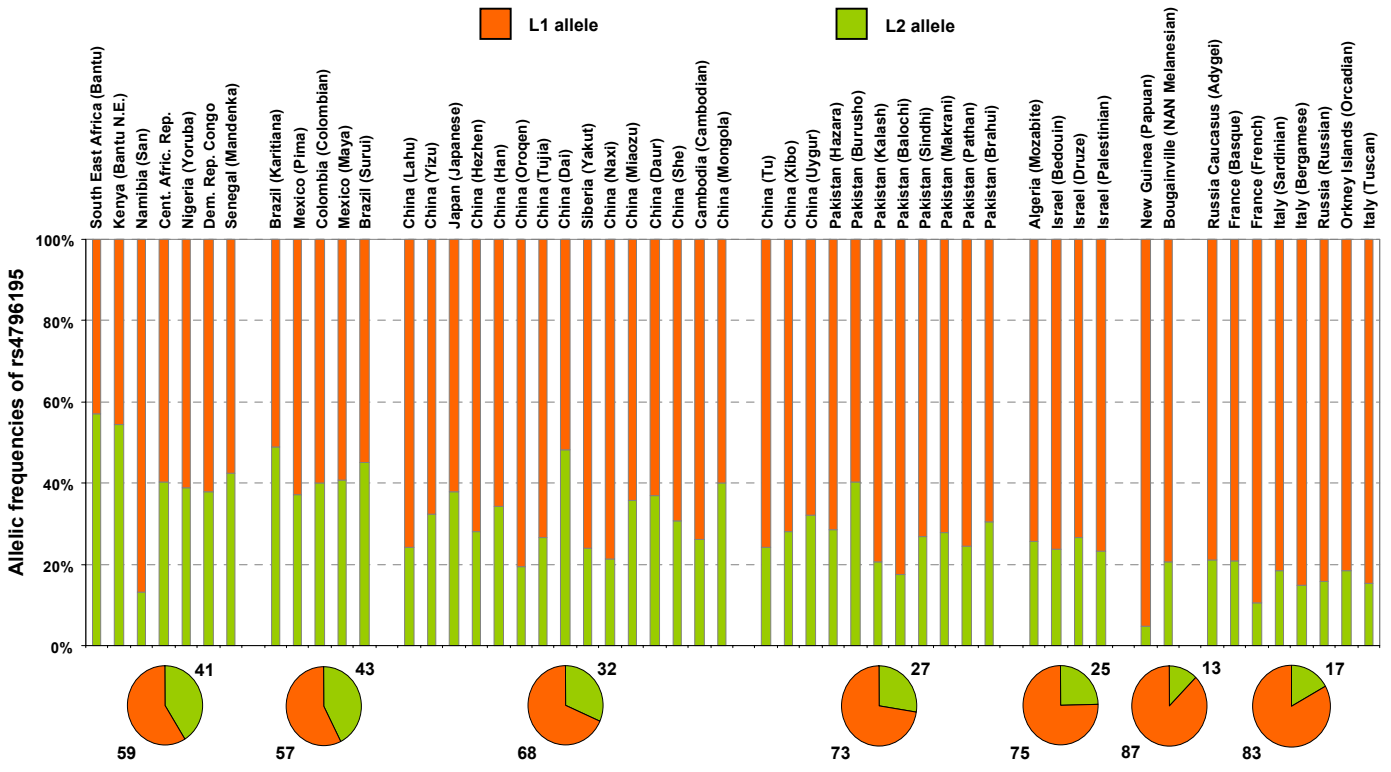


Figure 2

A



B

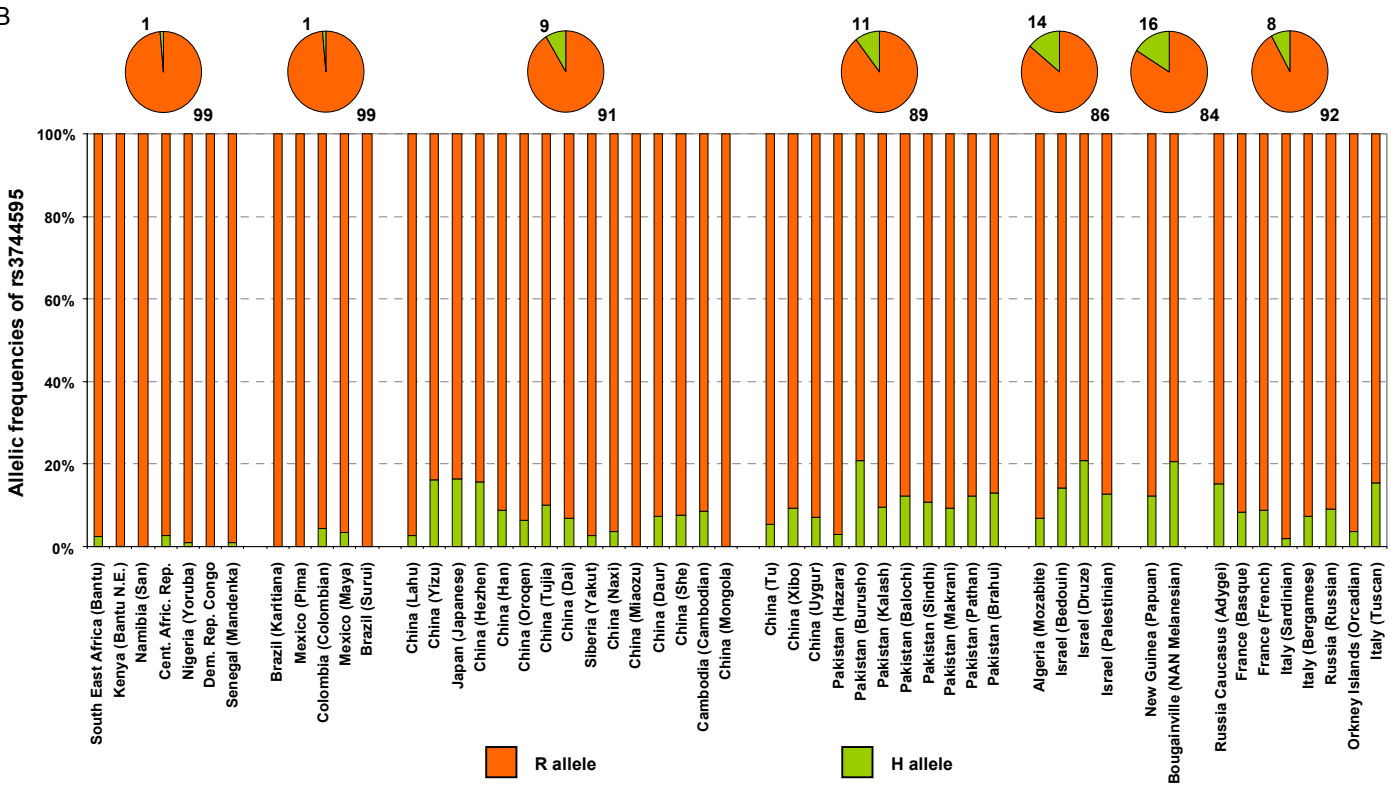


Figure 3

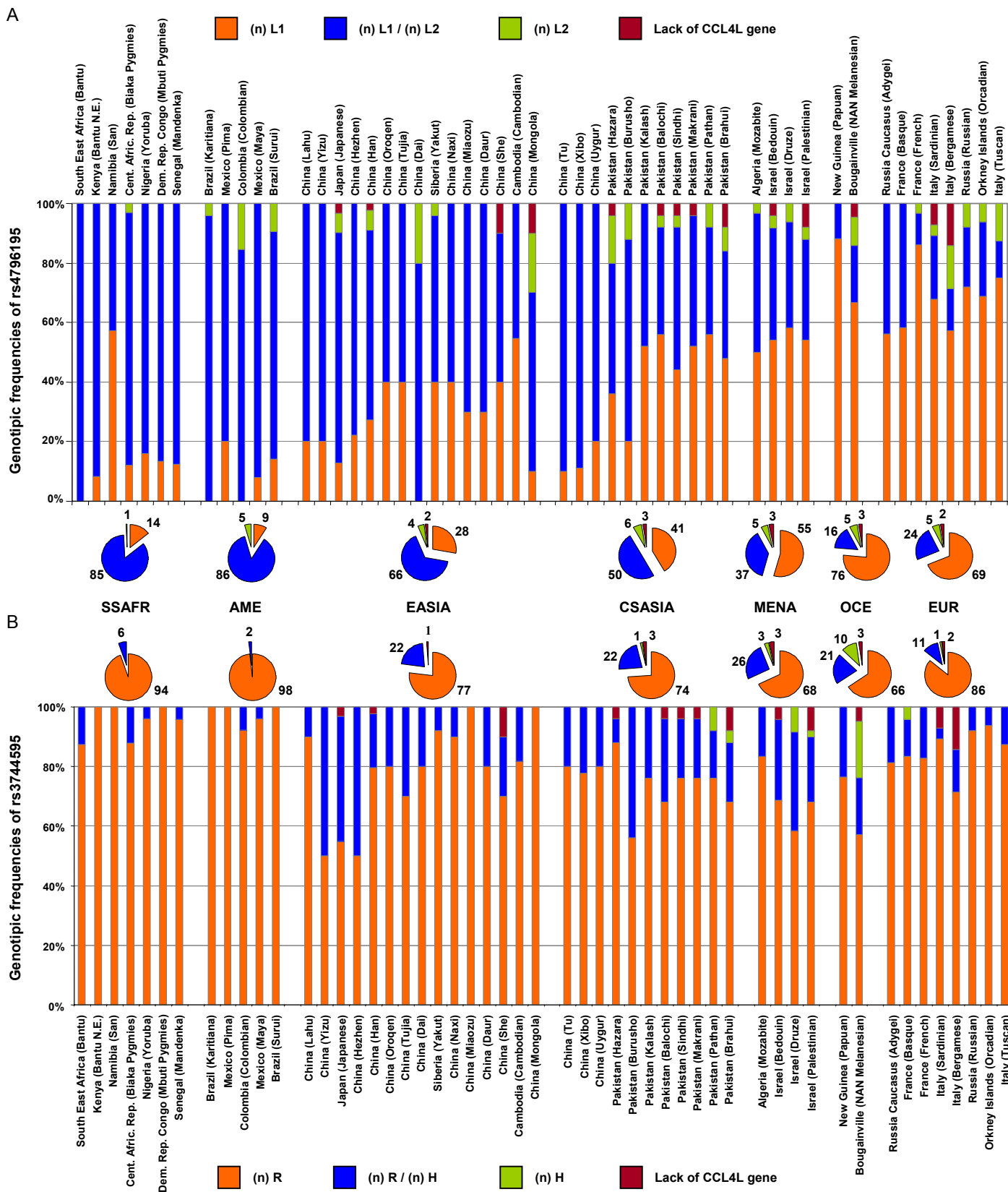


Figure 4

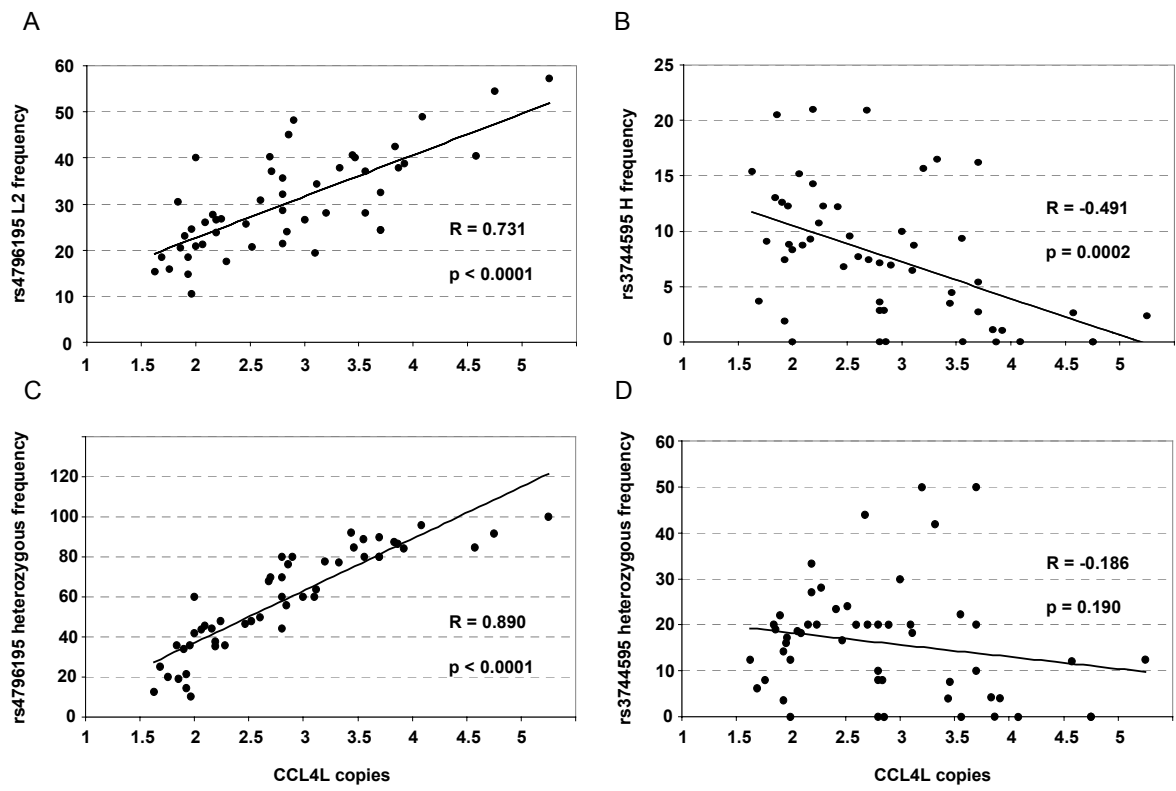


Figure 5



## 4.4. CAPÍTOL IV

### OBJECTIU:

Realitzar una mapa exhaustiu de l'organització genòmica dels membres de la superfamília de les quimiocines i establir connexions entre aquesta organització genòmica i les seves funcions.

### ARTICLE:

**The chemokine network. I. How the genomic organization of chemokines contains clues for deciphering their functional complexity.**

R. Colobran, R. Pujol-Borrell, M<sup>a</sup> P. Armengol and M. Juan.

*Clinical and Experimental Immunology*, **2007**, 148: 208–217.



#### 4.4.1. CONSIDERACIONS PRÈVIES SOBRE L'ARTICLE

Un dels aspectes que vam tenir en compte a l'inici de la meua etapa pre-doctoral va ser el de no centrar excessivament meua tasca de recerca a aspectes molt concrets del sistema immunitari. La idea era combinar la investigació enmarcada dins el projecte de recerca específic (en el nostre cas, l'estudi de la diversitat de les quimiocines CCL4 i CCL4L) amb la formació general en el camp de la immunologia, amb un èmfasi preferencial en les quimiocines en general. Per tant, tot el procés d'investigació de les quimiocines CCL4 i CCL4L, es combinava amb l'estudi global sobre la superfamília de les quimiocines. El biaix genètic que va anar prenent el projecte, també va influir en aquesta visió general que preteníem tenir sobre les quimiocines i el seu paper en la resposta immune i, després d'exhaustives búsquedes bibliogràfiques, vam adonar-nos de la manca de publicacions que estudiessin les quimiocines intentant establir una relació entre els aspectes genètics i funcionals. En aquest punt van començar a gestar-se les dues revisions que es presenten en aquest capítol i en el següent. Són dues revisions independents però que mantenen un vincle conceptual i, malgrat que es poden llegir perfectament per separat, l'ordre natural de lectura seria el que es presenta en aquesta tesi que, de fet, és l'ordre en que han estat publicades.

#### 4.4.2. RESUM DE DADES DEL QUART CAPÍTOL:

La primera revisió, que es presenta a continuació, pretén establir un lligam entre l'organització genòmica de les quimiocines i les seves múltiples funcions. Cal tenir en compte que la complexitat d'aquesta superfamília (amb més de 40 membres i una gran diversitat funcional) fa que sigui difícil iniciar-se en el seu coneixement. La nostra idea va ser (i és) que l'estudi conjunt de com s'estructuren els gens de les quimiocines i de les seves funcions facilita una visió més entenedora d'aquesta superfamília. Això és així per la història evolutiva d'aquestes molècules que, de forma general, s'estructuren en grups de gens (o *clusters*) que codifiquen proteïnes amb funcions relacionades. Aquest estudi presenta, a més a més, el primer mapa genètic complert que inclou tots els membres de la superfamília de les quimiocines humanes amb els seus respectius receptors.





# The chemokine network. I. How the genomic organization of chemokines contains clues for deciphering their functional complexity

R. Colobran, R. Pujol-Borrell,  
M<sup>a</sup> P. Armengol and M. Juan  
*Laboratory of Immunobiology for Research and  
Application to Diagnosis (LIRAD), Tissue and  
Blood Bank (BST), Institut d'Investigació en  
Ciències de la Salut Germans Trias i Pujol  
(IGTP), Badalona and Department of Cell  
Biology, Physiology and Immunology, Universitat  
Autònoma de Barcelona (UAB), Bellaterra,  
Barcelona, Spain*

Accepted for publication 18 January 2007  
Correspondence: Manel Juan Otero, Hospital  
Germans Trias i Pujol, Edifici de Recerca,  
Carretera de Can Ruti, cam de les escoles,  
s/n, 08916 Badalona, Spain.  
E-mail: mjuan.liradbst.germanstrias@  
gencat.met

## Chemokines: a special case of complexity

Chemokines are a large superfamily of small (approximately 8–15 kDa) structurally related cytokines that regulate cell trafficking of various types of leucocytes to areas of injury and play different roles in both inflammatory and homeostatic processes [1–3]. To date, 42 chemokines and 18 chemokine receptors (CKRs) have been identified in humans, but this number does not take into account several described variants that can increase their effective complexity.

Chemokines share from 20% to 95% of the amino acid sequence identity and have been grouped into four subfamilies, CXC, CC, CX3C and C chemokines, according to their structural cysteine motif found near the NH<sub>2</sub> terminus [4]. All of them adopt a similar folding pattern even in cases of

## Summary

Chemokines are a superfamily of small structurally related cytokines that have evolved to form a complex network of proteins that typically regulate leucocyte traffic but also carry very diverse sets of immune and non-immune functions. Two general features of cytokines, redundancy and promiscuity, are particularly prominent in chemokines. In part, these properties result from repeated processes of gene duplication and diversification, which has led to the present complex genomic map of chemokines, which contains cases of non-allelic isoforms, copy number polymorphisms and classical allelic variation. This genomic complexity is compounded with pre-translational and post-translational mechanisms resulting in a complex network of proteins whose essential functions are maintained, constituting a remarkable case of robustness reminiscent of crucial metabolic pathways. This reflects the adaptation of a system under strong evolutive pressure, supporting the concept that the chemokine system is essential for the coordination, regulation and fine-tuning of the type of immune response. In this first review, we analyse currently available data on the chemokine superfamily, focusing on its complex genomic organization. Genes encoding essential inflammatory chemokines are grouped into defined chromosomal locations as clusters and miniclusters that, from the genetic point of view, can be considered single entities given their overall functions (many ligands of a cluster bind to a few shared receptors). We will try to interpret this genomic organization of chemokines in relation to the main functions acquired by each individual member or by each cluster. In a second review, we shall focus on the relationship of chemokine variability and disease susceptibility.

**Keywords:** chemokines, evolution, genome, human, variability

low overall sequence identity: the chemokine scaffold consists of a N-terminal region followed and connected via cysteine bonds to its core of three  $\beta$ -sheets and a C-terminal  $\alpha$ -helix. Chemokines carry out their biological functions through binding to seven-transmembrane domain G protein-coupled receptors expressed on different leucocytes, endothelial and other cell types [5–7].

Based upon the site and circumstance of their expression, chemokines can generally be classified as either inducible (inflammatory) or constitutive (homeostatic). Inducible chemokines are usually expressed under inflammatory stimuli. In contrast, constitutive chemokines are expressed in the absence of infection or damage, controlling from cell trafficking in the embryo and foetus to homeostatic leucocyte homing for immune surveillance.

In evolution, diversification through the generation of multiple alleles is very common and the immune system contains several groups of genes with prominent allelic variation, the MHC genes being the most extreme example in the human genome. The chemokine system constitutes a less well-studied but very revealing case of how through evolution, a complex network of genes has acquired a very diverse set of related functions. In this review, we analyse the existing correlations between the genomic organization of the different groups of chemokines as clusters and miniclusters and their linked function as we believe that this perspective will help to better understand the sometimes confusing chemokine system.

### **How the chemokine network conciliates fine-tuning, reliability and robustness with extreme variability and phylogenic plasticity, by using redundancy and promiscuity**

To preserve the overall function, the chemokine network is relatively insensitive to alterations of their individual components. In fact, the chemokine superfamily has achieved, in spite of its broad variability, a high degree of robustness, mainly through the redundancy and binding promiscuity of their ligands and receptors [8]. This type of complexity is a general feature of cytokines, but reaches its highest level in the chemokine superfamily. Chemokines are redundant in their effects on the target cells (no chemokine is uniquely active on one leucocyte population and usually a given leucocyte population has receptors for different chemokines), and the interaction with their receptors also has considerable promiscuity (most known receptors interact with multiple ligands and most ligands interact with more than one receptor).

Although redundancy among chemokines is not universal, there is a considerable degree of functional overlapping among some of them (this also applies to CKRs). The chemokine/CKR system robustness makes it possible that when one given chemokine or receptor is defective there is usually an alternative set of chemokines or CKRs that can maintain the main biological functions. This principle has a striking exception: the CXCL12/CXCR4 pair. CXCL12 and CXCR4-knockout mice show embryonic lethality [9,10], providing a clear example of the wide range of biological effects of chemokines, whose functions extend beyond simply attracting leucocytes, taking part in organ development, angiogenesis, angiostasis and immune regulation.

In spite of the fact that one chemokine might act on different cells and, conversely, a single cell responds to many chemokines, the role of chemokines in both inflammatory and homeostatic processes is finely regulated case-by-case by the secretion of specific sets of chemokines and by a spatial and temporal control of chemokine production (as well as regulation of receptor expression on particular leucocyte subtypes). The comparison of the properties of each of the

different ligands of a common receptor suggests the existence of additional mechanisms to limit redundancy [11]. Two ligands that share a given receptor may exhibit significant differences in their affinity for the receptor, their ability to bind to glycosaminoglycans (GAGs, see below), their capacity to induce receptor desensitization, their half-life, their susceptibility to protease modifications, or in the territories of distribution (e.g. extracellular matrix, endothelium, etc.). A new level of chemokine regulation has emerged from the discovery that chemokines can also act as receptor antagonists. By this dual-activity, chemokines would attract or activate a population of cells bearing one receptor, while at the same time preventing the recruitment and/or activation of a different cell population via another receptor [12].

Another important mechanism to control the chemokine activity through regulation of chemokine levels in the body is carried out by decoy receptors, defined as cell surface ligand-binding proteins with high affinity and specificity for chemokines, but which are unable to induce downstream signalling [13]. Three of these so-called 'silent' receptors have been identified, DARC [14], D6 [15–17] and CCX-CKR [18,19], each of them interacting with a specific set of chemokines. This chemokine scavenging is important for maintaining chemokine homeostasis because efficient recruitment of inflammatory cells requires low chemokine levels in the bloodstream and high chemokine levels in the target tissue.

In addition to chemokine G protein-coupled receptor interactions, chemokines bind both to soluble glycosaminoglycans (GAGs) and GAGs immobilized on cell surfaces and the extracellular matrix. Chemokine immobilization through the GAG interaction facilitates the formation of haptotactic chemokine gradients and enhances their concentration at the site of production [20]. Interestingly, recent *in vivo* work exploiting recombinant chemokine mutants suggests that oligomeric chemokine binding to GAGs is crucial for biological responses [21]. Interaction with GAGs may also provide another level of specificity and control to cell migration, beyond that defined by receptor engagement, by selective binding of certain chemokines to different types of GAGs.

In summary, all these mechanisms seem to operate to increase the selectivity of cell recruitment and, in more general terms, to provide mechanisms to exert a fine control of the variability, redundancy and promiscuity of the chemokine network.

## **Genomic organization of chemokines**

### **Genomic evolution, the first source of variability**

Chemokine function probably preceded the origin of the chemokine network as chemokine-like molecules have been detected in sponges [22]. Numerous studies have pointed

out that most, if not all, chemokines arose by gene duplication of a single ancestral gene. In fact, chemokine and CKR evolution can be traced through phylogeny from early vertebrates to non-human primates [23] and, overall, chemokines have expanded markedly their role in orchestrating the immune response and in organizing the lymphoid tissue. Co-evolution of pathogens with their hosts has led to adaptive changes where some pathogens, e.g. viruses, encode chemokine homologues as part of their evasion strategy [24,25].

During the evolution, the different options for gene variability (from genomic region duplications to point mutations) have produced the present genomic organization of chemokines in humans [26]. Many chemokines (just as many chemokine receptors) are clustered in defined chromosomal locations. Two main clusters have been recognized, both of them codifying the essential inflammatory chemokines: the CXC cluster, located in chromosome 4q12–21, and the CC cluster, located in chromosome 17q11.2 (Fig. 1). The chemokines that map in the CXC and the CC clusters seem to maintain some specific functions: CXC cluster chemokines recruit mainly neutrophils while CC cluster members typically attract mononuclear cells.

Genes of the more recently identified CC and CXC chemokines tend to be located in different chromosomal locations, far from CC and CXC clusters. These chemokine genes located away from the two major clusters correspond to older genes in evolutionary terms, remaining better conserved among species probably because of their very specific functions (in contrast, the major CXC and CC chemokine clusters were generated more recently). An important characteristic of chemokine genes from the same cluster is that they code for many ligands that interact with a few receptors. Therefore, chemokine clusters are single entities based on their overall function [4] (Fig. 2). For this reason, in this review we have used the genomic organization as the guidance for discussing the chemokine system.

### Human CXC chemokines subfamily

The human CXC chemokines subfamily comprises 15 ligands and six receptors in humans, and their functions extend well beyond the immune system. Most CXC chemokines are evolutionary recent and exclusive of mammals: only CXCL12 (SDF-1) and CXCL14 (BRAK), two homeostatic chemokines, have unambiguous orthologues in fish, suggesting that they are the phylogenetically modern representatives of the ancestral CXC chemokine [27].

Almost all the CXC inflammatory chemokines are located in the chromosome 4 cluster, organized into two subclusters or groups.

**1** The centromeric subcluster contains the genes coding for CXCL1 to CXCL8 (systematic names for GRO $\alpha$ , GRO $\beta$ , GRO $\gamma$ , PF4, ENA-78, GCP-2, NAP-2 and IL-8). All

members of this group, with the exception of CXCL4, show the tri-peptide ELR (glutamic-leucine-arginine) motif (next to the first conserved cysteine) implicated in neutrophil chemotaxis through binding to CXCR1 and CXCR2, and also conferring on them pro-angiogenic effects. Despite their similarities, each ELR+ chemokine has distinct effects, due in part to their divergent abilities to bind and activate CXCR1 and CXCR2. The redundancy in ELR+ CXC chemokines and their receptors may provide multiple levels of regulation that allow for specific control of inflammation [28]. Completely different from the other chemokines of this group, CXCL4 is an anti-angiogenic ELR-negative chemokine. CXCL4 (mainly found in platelets) inhibits endothelial cell proliferation/migration and angiogenesis. Growing evidence suggests that CXCL4 biology depends on its unusually high affinity for heparan sulphates, rather than the binding to a well-defined CXCR [29–31]. However, recently, CXCR3-B (an alternatively spliced variant of CXCR3) has been identified as a CXCL4 receptor [32]. Besides, CXCL4 has a non-allelic copy named CXCL4L1 (PF4V1) whose product, although it only differs by three amino acids, has a more potent anti-angiogenic factor [33,34], making it a very interesting molecule for therapy [35].

**2** The telomeric subcluster of inflammatory CXC chemokines in chromosome 4 includes ELR-negative chemokines CXCL9 (Mig), CXCL10 (IP-10) and CXCL11 (I-TAC). These three chemokines bind to CXCR3 (CXCR3-A and CXCR3-B) and attract activated Th1 lymphocytes and NK cells. In contrast to ELR+ chemokines, these three chemokines are potent angiostatic factors. Therefore the CXC chemokines are a unique family of cytokines due to their ability to behave in an opposite manner in the regulation of angiogenesis through the ELR motif and this different activity may have an impact on the pathogenesis of a variety of inflammatory disorders [36]. Additionally, CXCL9, CXCL10 and CXCL11 are natural antagonists for CCR3, suggesting that chemokines that attract Th1 cells via CXCR3 can concomitantly block the migration of Th2 cells in response to CCR3 ligands, thus enhancing the polarization of T cell recruitment [37]. The recently discovered CXCR3 receptor variant CXCR3-B (in addition to the classic form, CXCR3-A) could explain the different functions of CXCL9, CXCL10 and CXCL11 on distinct cell types.

Aside from these two subclusters of CXC inflammatory chemokines but in the same stretch of chromosome 4, we find CXCL13, a homeostatic CXC chemokine constitutively expressed in lymphoid follicles. It is the unique ligand of CXCR5, and essentially contributes to B cell homing and proper positioning of these cells within the microanatomic compartments of secondary lymphoid organs [38] and specifically in defining the B and T cell areas in the lymphoid follicles.

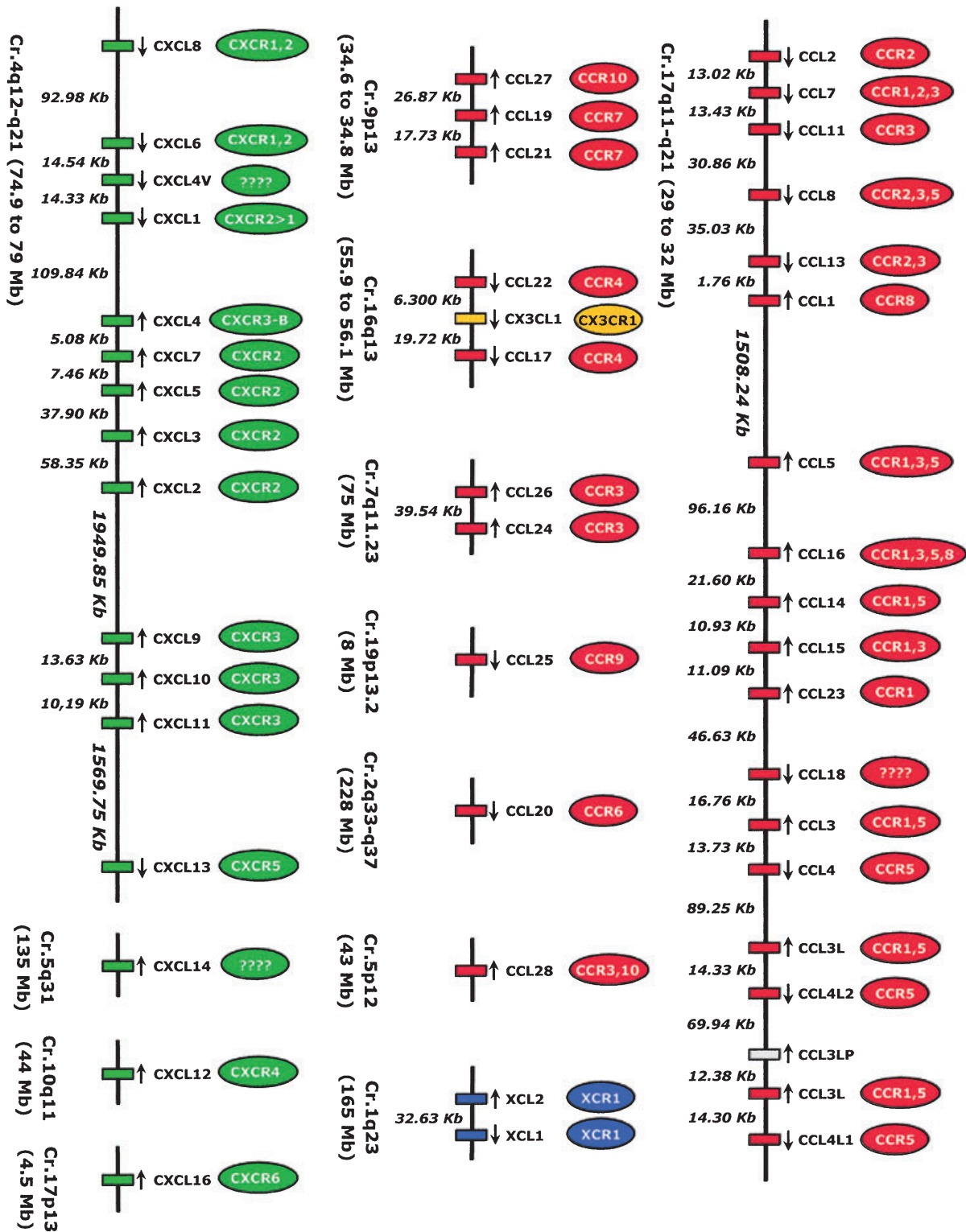
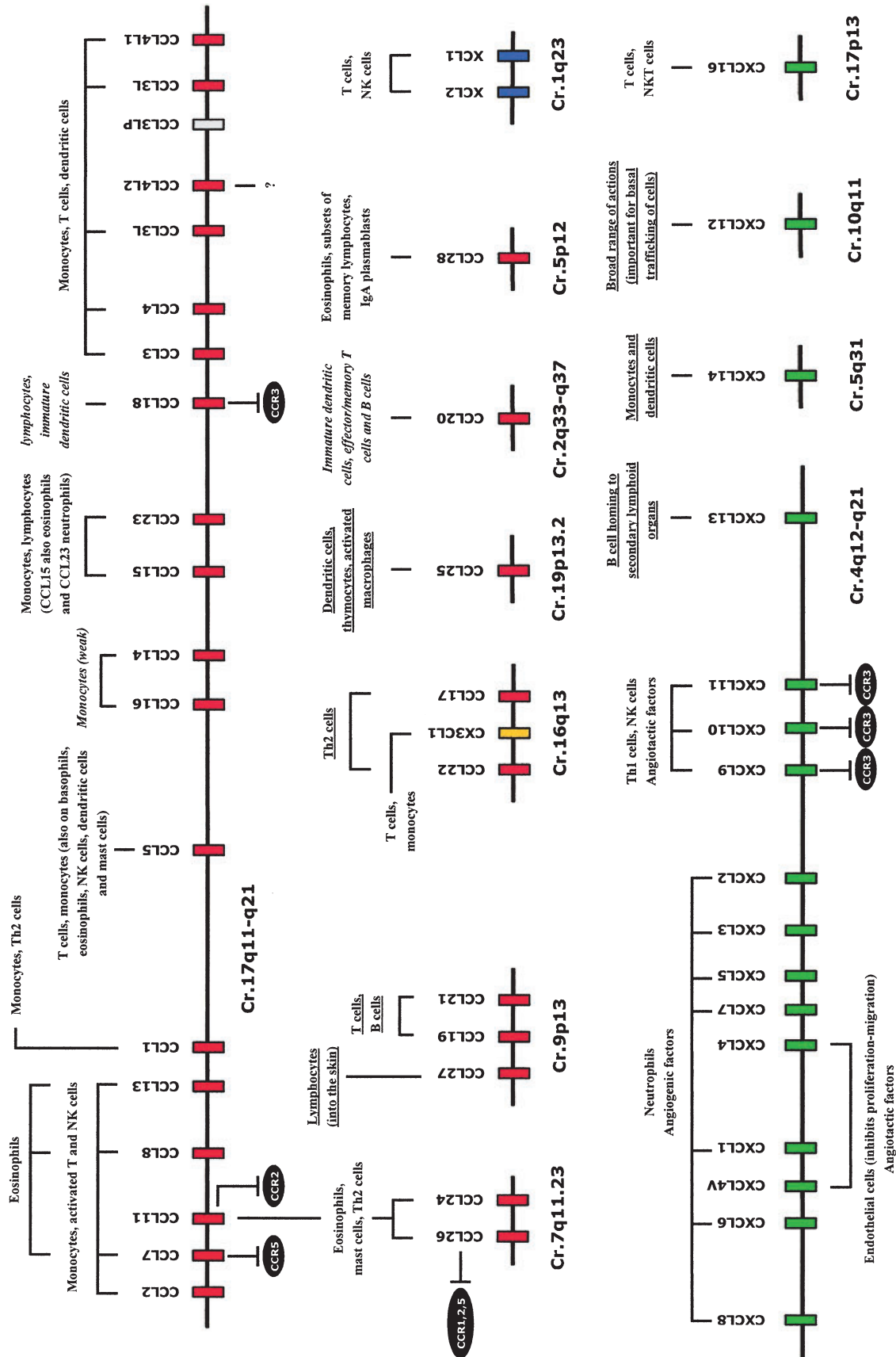


Fig. 1. Map of genomic organization of human chemokines. CC chemokines in red, CXC chemokines in green, CX3C chemokine in yellow and C chemokines in blue. Distances between genes are expressed in Kb. Gene or cluster chromosomal location is expressed as a distance from the beginning of chromosome (in Mb). Receptors are shown above each ligand. The orientation of each gene is shown by an arrow.



**Fig. 2.** Relationship between genomic organization and function of chemokines. Main cell types (and essential actions) targeted by individual or chemokine cluster are shown. Functions of homeostatic chemokines are underlined and functions of homeostatic/inducible chemokines are in italics. CC chemokines in green, CXC chemokines in green, CX3C chemokine in yellow and C chemokines in blue. Antagonist functions are represented by ⊥.

The three remaining CXC chemokines are mapped separately in other chromosomes. The already mentioned CXCL12 (SDF-1, chromosome 10) is a homeostatic CXC chemokine widely expressed and showing a broad range of actions, affecting immune cell chemotaxis to neural development. CXCL12 appears to be particularly important for the regulation of homeostatic traffic and distribution of cells in the different compartments and subcompartments of the immune system, for instance directing naïve T cell traffic through lymph nodes [39,40]. CXCL12 is important for HIV infection because CXCR4, its only known receptor, is also the membrane protein used by HIV T-tropic strains as a co-receptor to entry into target cells [41–43].

CXCL14 (BRAK, chromosome 5) is also a homeostatic CXC chemokine ubiquitously expressed in all normal tissues. The structure of CXCL14 is considerably divergent from all other chemokines, its receptor has not been identified and its activity is not yet well defined. However, it seems that CXCL14 is chemoattractant for monocytes [44] and dendritic cells [45,46] and recent data indicate that CXCL14 expression also inhibits tumour growth [47].

The last CXC chemokine, CXCL16 (chromosome 17) is, in the strictest definition, a member of the CXC family but it is distantly related to all known chemokines and might be phylogenetically closer to CC chemokines. Additionally, its structure is similar to CX3CL1 in having a transmembrane region and a chemokine domain suspended by a mucin-like stalk [48,49]. Thus, CXCL16 not only attracts T cells and NKT cells toward DCs but also supports their firm adhesion to DCs [50]. Interestingly, after proteolytic cleavage by members of the ADAM family of proteases, the chemokine domain of CXCL16 can be released, attracting cells expressing the same receptor [51,52].

### Human CC chemokines subfamily (and CX3C)

Within the cluster of pro-inflammatory CC chemokines in chromosome 17, we clearly distinguish two subclusters:

**1** Centromerically, we found CCL2, CCL7, CCL11, CCL8, CCL13 and CCL1 that constitute the 'MCP' subcluster. CCL2, CCL8, CCL7 and CCL13 (originally called MCP-1, -2, -3 and -4) are closely related chemoattractants for monocytes, activated T and NK cells; additionally CCL8, CCL7 and CCL13 also attract eosinophils. They bind mainly to CCR2 and CCR3, and may be pathogenic in human diseases characterized by mononuclear and/or eosinophilic infiltration (as atherosclerosis and asthma). Interestingly, CCL7 is a highly potent CCR5 antagonist [53]. CCL11 (eotaxin-1), a potent eosinophil chemoattractant, maps within the MCP cluster, although the two other members of the eotaxin gene family, CCL24 (eotaxin-2) and CCL26 (eotaxin-3), are located together in a minicluster in chromosome 7. These three chemokines recruit and activate CCR3-bearing cells such as eosi-

nophils, mast cells and Th2 lymphocytes, playing a major role in allergic disorders. Interestingly, CCL26 is the first human chemokine that features broad antagonistic activities as a natural antagonist to the ligands of CCR1, CCR2 and CCR5 (CCL11 can also block CCR2 [54]). This suggests that CCL26 may have a modulator rather than an inflammatory function [12,55]. Finally, the last and more telomeric gene codes for CCL1 (I-309). CCL1 has structural features that distinguish it from other chemokines, including an additional pair of cysteines that form a third intramolecular disulphide bond, and a propensity to remain monomeric at concentrations at which many other chemokines dimerize. CCL1 is the only high-affinity ligand for CCR8 and in spite of these peculiarities it has similar functional features to the other members of this subcluster, being also chemotactic for monocytes and Th2 lymphocytes [56].

**2** The second subcluster of pro-inflammatory CC chemokines in chromosome 17 is 1.5 Mb telomeric to the MCP one. To simplify, in this subcluster we find CCR1/CCR5 ligands that we can in turn subdivide into the HCC and MIP groups. (a) The HCC group contains CCL14, CCL15, CCL16 (the systematic names of HCC1, HCC2, HCC4), CCL23 (MPIF) and CCL5 (RANTES). CCL14, CCL15, CCL16 and CCL23 are clustered within a 40-kb long region whereas CCL5 (a very promiscuous chemokine, binding with high affinity to CCR1/CCR3/CCR5) is located approximately 100 Kb away (with several non-related genes within). The HCC genes are relatively large, from 3.1 Kb (CCL14) to 8.8 kb (CCL5), compared with other CC chemokine genes, such as CCL4 and CCL2 (< 2 kb). Moreover, in contrast to most other human CC chemokine genes that comprise three exons, the CCL15 and CCL23 have four exons. In fact, their high nucleotide homology indicates that they have been generated recently by duplication [57]. All these chemokines share the binding to CCR1, even though most of them can bind to other CCRs. This group is functionally more heterogeneous (see Fig. 2) than the MIP group or the MCP subcluster [58–63]. (b) The MIP group is composed of CCL18 (PARC), CCL3 (MIP-1 $\alpha$ ), CCL3L (LD78 $\beta$ ), CCL4 (MIP-1 $\beta$ ) and CCL4L. This is a very interesting gene minicluster that could serve as a model of chemokine evolution. CCL3 and CCL4 were formed by duplication of a common ancestral gene [64] and both have a second non-allelic copy, CCL3L and CCL4L, originated by a second duplication of the entire DNA block that contains the two genes. Next to them, can also be found a CCL3 5'-truncated pseudogene, named CCL3LP (LD78 $\gamma$ ) [65]. The CCL3L-CCL4L tandem shows also an additional variability: they have a copy number polymorphism (CNP) that means that these genes are present in variable numbers in the genome. Regarding this CNP, there are not only interindividual and interpopulation differences in the copy number of the CCL3L-CCL4L tandem, but,

interestingly, there are also haplotypes with different numbers of individual genes [66]. Moreover, a recent study demonstrated that there are two main allelic variants of CCL4L: CCL4L1, the originally described variant, and CCL4L2, in which one base substitution at the acceptor splice site of intron 2 originates a new complex pattern of splicing variants [67]. Functionally, CCL3/CCL3L and CCL4/CCL4L are highly related. They both bind to CCR5, but CCL3 binds also to CCR1 with high affinity [64,68].

The other member of this group and the most centromeric one, CCL18, is larger than average (7.2 kb *versus* general size of 2–3 kb) as a result of a first intron of 6.0 kb, which contains two pseudo-exons. This feature, together with the presence of two adjacent regions within the CCL18 gene sharing high-sequence similarity with the CCL3 gene, suggests that the CCL18 gene may have been generated by fusion of two CCL3-like genes with deletion and selective use of some exons [69]. CCL18 apparently functions as an inflammatory/inducible or constitutive/homeostatic chemokine depending on the circumstances: it is mainly expressed by a broad range of monocytes/macrophages and DC, and it is constitutively present at high levels in human plasma (probably contributing to the physiological homing of lymphocytes and DC and to the generation of primary immune responses) [70]. Importantly, the CCL18 receptor has not been yet identified. This chemokine exhibits CCR3 antagonistic activity [71].

The rest of the CC chemokines are located in other chromosomes either as a single member or constituting a minicuster. CCL27 (CTACK), CCL19 (ELC) and CCL21 (SLC) are grouped in a minicuster in chromosome 9. CCL19 and CCL21 share only 32% amino acid identity but constitute a genetically and functionally highly related subgroup of homeostatic CC chemokines. They are the only two ligands of CCR7, and they have a potent chemotactic activity for naive T cells and B cells. They clearly play a pivotal role in naive lymphocyte homing and traffic within lymphoid tissues [72,73]. Unrelated to CCL19 and CCL21, CCL27 mediates the migration of lymphocytes into the skin, through binding to the CCR10 [74].

The two closely related chemokines CCL22 (MDC) and CCL17 (TARC) map in chromosome 16 together with CX3CL1 (Fractalkine), the only member of CX3C subfamily. CCL17 and CCL22, the two known ligands of CCR4, are a pair CC chemokines highly expressed in the thymus. It is known that they selectively attract Th2 type memory T cells into the inflammatory sites and regulate Th2-related immune responses [75,76]. Interestingly, the CX3CL1 gene is situated between the CCL17 and CCL22 genes. Although the three-dimensional structure of the chemokine domain of CX3CL1 is different from that of CC chemokines, it was identified thanks to its amino acid homology with members of the CC chemokine family [77]. Unlike other chemokines, CX3CL1 carries the chemokine domain on top of an

extended mucin-like stalk and can exist in two forms: as a transmembrane chemokine/mucin hybrid protein and as a soluble chemotactic polypeptide generated by proteolytic cleavage by members of the ADAM family of proteases [78].

CCL20 (LARC, chromosome 2) is the only chemokine known to interact with CCR6, a property shared with the anti-microbial peptides  $\beta$ -defensins [79]. Like CCL18, CCL20 is considered a constitutive and inducible chemokine. Constitutive expression of CCL20 was originally demonstrated in the liver, mucosa-associated and lymphoid tissues and induction in monocytes as well as many other cells [80].

CCL25 (TECK, chromosome 19) is the only chemokine binding to CCR9. It is constitutively and selectively expressed at high levels in the thymus and small intestine. CCL25 has been reported to chemoattract dendritic cells, thymocytes and activated macrophages and probably plays a significant role in the recruitment of developing thymocytes to discrete compartments within the thymus and, in general, in T lymphocyte development [81,82].

The last CC chemokine, CCL28 (MEC, chromosome 5), is most homologous to CCL27, displaying about 40% identity and sharing binding to CCR10 although CCL28 also binds to CCR3. It is expressed not in the skin as CCL27 but, instead in diverse mucosal tissues, suggesting that CCL28 may play an important role in the physiology and/or recruitment of specialized cells into mucosal tissues [83–85].

### Human C chemokines subfamily

The C chemokines subfamily, located in a minicuster in chromosome 1, has only two highly homologous genes encoding XCL1 (lymphotactin- $\alpha$ ) and XCL2 (lymphotactin- $\beta$ ). These two chemokines that differ by only two amino acid residues both bind to XCR1, and have some homology with a number of CC chemokines, especially with CCL8 and CCL3. The major structural feature of XCL1/2 is that they lack two of the four-cysteine residues characteristics of the chemokines (and therefore have only one disulphide bond). XCL1/2 are potent chemoattractants of T and NK cells, but not of monocytes or neutrophils, and they are the major products of activated CD8+ T cells [86,87].

### Concluding remarks

The human chemokine superfamily forms a complex and robust network of genes and proteins playing very diverse roles but also displaying a lot of functional interactions between them. Current genomic organization of human chemokine genes provides a useful tool to better understand the complexity of this superfamily. Based on an exhaustive genomic map of all human chemokines, we have linked the genetic organization (clusters and miniclusters) with the main functions, focusing on inflammatory chemokines. Future studies on the relationship between genetics and function should focus on other important mechanisms of



variability, such as polymorphisms and alternative splicings, as they may lead to disease susceptibility. In this context, it would be important not only to deal with each chemokine separately but also to take into account all genetically and functionally related members (a second review will mainly focus on these aspects).

## Acknowledgements

This work was supported by grants from the FIPSE (Fundación para la Investigación y la Prevención del Sida en España) (project 36487/05) and FIS (Fondo de Investigaciones Sanitarias) (project PI 02/0104).

## References

- Springer TA. Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. *Cell* 1994; **76**:301–14.
- Luster AD. Chemokines – chemotactic cytokines that mediate inflammation. *N Engl J Med* 1998; **338**:436–45.
- Baggiolini M. Chemokines and leukocyte traffic. *Nature* 1998; **392**:565–8.
- Zlotnik A, Yoshie O. Chemokines: a new classification system and their role in immunity. *Immunity* 2000; **12**:121–7.
- Murphy PM. The molecular biology of leukocyte chemoattractant receptors. *Annu Rev Immunol* 1994; **12**:593–633.
- Premack BA, Schall TJ. Chemokine receptors: gateways to inflammation and infection. *Nat Med* 1996; **2**:1174–8.
- Zlotnik A, Morales J, Hedrick JA. Recent advances in chemokines and chemokine receptors. *Crit Rev Immunol* 1999; **19**:1–47.
- Mantovani A. The chemokine system: redundancy for robust outputs. *Immunol Today* 1999; **20**:254–7.
- Tachibana K, Hirota S, Iizasa H *et al.* The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature* 1998; **393**:591–4.
- Ma Q, Jones D, Borghesani PR *et al.* Impaired B-lymphopoiesis, myelopoiesis, and derailed cerebellar neuron migration in CXCR4- and SDF-1-deficient mice. *Proc Natl Acad Sci USA* 1998; **95**:9448–53.
- Devalaraja MN, Richmond A. Multiple chemotactic factors: fine control or redundancy? *Trends Pharmacol Sci* 1999; **20**:151–6.
- Petkovic V, Moghini C, Paoletti S, Ugucioni M, Gerber B. Eotaxin-3/CCL26 is a natural antagonist for CC chemokine receptors 1 and 5. A human chemokine with a regulatory role. *J Biol Chem* 2004; **279**:23357–63.
- Comerford I, Nibbs RJ. Post-translational control of chemokines: a role for decoy receptors? *Immunol Lett* 2005; **96**:163–74.
- Lee JS, Frevert CW, Wurfel MM *et al.* Duffy antigen facilitates movement of chemokine across the endothelium in vitro and promotes neutrophil transmigration in vitro and in vivo. *J Immunol* 2003; **170**:5244–51.
- Fra AM, Locati M, Otero K *et al.* Cutting edge: scavenging of inflammatory CC chemokines by the promiscuous putatively silent chemokine receptor D6. *J Immunol* 2003; **170**:2279–82.
- Weber M, Blair E, Simpson CV *et al.* The chemokine receptor D6 constitutively traffics to and from the cell surface to internalize and degrade chemokines. *Mol Biol Cell* 2004; **15**:2492–508.
- Galliera E, Jala VR, Trent JO *et al.* beta-Arrestin-dependent constitutive internalization of the human chemokine decoy receptor D6. *J Biol Chem* 2004; **279**:25590–7.
- Gosling J, Dairaghi DJ, Wang Y *et al.* Cutting edge: identification of a novel chemokine receptor that binds dendritic cell- and T cell-active chemokines including ELC, SLC, and TECK. *J Immunol* 2000; **164**:2851–6.
- Townson JR, Nibbs RJ. Characterization of mouse CCX-CKR, a receptor for the lymphocyte-attracting chemokines TECK/mCCL25, SLC/mCCL21 and MIP-3beta/mCCL19: comparison to human CCX-CKR. *Eur J Immunol* 2002; **32**:1230–41.
- Tanaka Y, Adams DH, Hubscher S, Hirano H, Siebenlist U, Shaw S. T-cell adhesion induced by proteoglycan-immobilized cytokine MIP-1 beta. *Nature* 1993; **361**:79–82.
- Proudfoot AE, Handel TM, Johnson Z *et al.* Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. *Proc Natl Acad Sci USA* 2003; **100**:1885–90.
- Muller WE, Wiens M, Muller IM, Schroder HC. The chemokine networks in sponges: potential roles in morphogenesis, immunity and stem cell formation. *Prog Mol Subcell Biol* 2004; **34**:103–43.
- Clay CC, Rodrigues DS, Brignolo LL *et al.* Chemokine networks and in vivo T-lymphocyte trafficking in nonhuman primates. *J Immunol Meth* 2004; **293**:23–42.
- McFadden G, Murphy PM. Host-related immunomodulators encoded by poxviruses and herpesviruses. *Curr Opin Microbiol* 2000; **3**:371–8.
- Boomker JM, de Leij LF, The TH, Harmsen MC. Viral chemokine-modulatory proteins: tools and targets. *Cytokine Growth Factor Rev* 2005; **16**:91–103.
- DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ. Defining the origins and evolution of the chemokine/chemokine receptor system. *J Immunol* 2006; **176**:401–15.
- Huisling MO, Stet RJ, Kruiswijk CP, Savelkoul HF, Lidy Verburg-van Kemenade BM. Molecular evolution of CXC chemokines: extant CXC chemokines originate from the CNS. *Trends Immunol* 2003; **24**:307–13.
- Bizzarri C, Beccari AR, Bertini R, Cavicchia MR, Giorgini S, Allegretti M. ELR+ CXC chemokines and their receptors (CXC chemokine receptor 1 and CXC chemokine receptor 2) as new therapeutic targets. *Pharmacol Ther* 2006; **112**:139–49.
- Sachais BS, Higazi AA, Cines DB, Poncz M, Kowalska MA. Interactions of platelet factor 4 with the vessel wall. *Semin Thromb Hemost* 2004; **30**:351–8.
- Bikfalvi A. Platelet factor 4: an inhibitor of angiogenesis. *Semin Thromb Hemost* 2004; **30**:379–85.
- Mixon TA, Dehmer GJ. Recombinant platelet factor 4 for heparin neutralization. *Semin Thromb Hemost* 2004; **30**:369–77.
- Lasagni L, Francalanci M, Annunziato F *et al.* An alternatively spliced variant of CXCR3 mediates the inhibition of endothelial cell growth induced by IP-10, Mig, and I-TAC, and acts as functional receptor for platelet factor 4. *J Exp Med* 2003; **197**:1537–49.
- Green CJ, Charles RS, Edwards BF, Johnson PH. Identification and characterization of PF4varl, a human gene variant of platelet factor 4. *Mol Cell Biol* 1989; **9**:1445–51.
- Eisman R, Surrey S, Ramachandran B, Schwartz E, Poncz M. Structural and functional comparison of the genes for human platelet factor 4 and PF4alt. *Blood* 1990; **76**:336–44.
- Struyf S, Burdick MD, Proost P, Van Damme J, Strieter RM. Platelets release CXCL4L1, a nonallelic variant of the chemokine platelet factor-4/CXCL4 and potent inhibitor of angiogenesis. *Circ Res* 2004; **95**:855–7.

- 36 Belperio JA, Keane MP, Arenberg DA *et al.* CXC chemokines in angiogenesis. *J Leukoc Biol* 2000; **68**:1–8.
- 37 Loetscher P, Pellegrino A, Gong JH *et al.* The ligands of CXC chemokine receptor 3, I-TAC, Mig, and IP10, are natural antagonists for CCR3. *J Biol Chem* 2001; **276**:2986–91.
- 38 Gunn MD, Ngo VN, Ansel KM, Eklund EH, Cyster JG, Williams LT. A B-cell-homing chemokine made in lymphoid follicles activates Burkitt's lymphoma receptor-1. *Nature* 1998; **391**:799–803.
- 39 Bleul CC, Fuhlbrigge RC, Casasnovas JM, Aiuti A, Springer TA. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J Exp Med* 1996; **184**:1101–9.
- 40 Aiuti A, Webb IJ, Bleul C, Springer T, Gutierrez-Ramos JC. The chemokine SDF-1 is a chemoattractant for human CD34+ hematopoietic progenitor cells and provides a new mechanism to explain the mobilization of CD34+ progenitors to peripheral blood. *J Exp Med* 1997; **185**:111–20.
- 41 Bleul CC, Farzan M, Choe H *et al.* The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* 1996; **382**:829–33.
- 42 Oberlin E, Amara A, Bachelier F *et al.* The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* 1996; **382**:833–5.
- 43 Moore JP, Trkola A, Dragic T. Co-receptors for HIV-1 entry. *Curr Opin Immunol* 1997; **9**:551–62.
- 44 Kurth I, Willmann K, Schaerli P, Hunziker T, Clark-Lewis I, Moser B. Monocyte selectivity and tissue localization suggests a role for breast and kidney-expressed chemokine (BRAK) in macrophage development. *J Exp Med* 2001; **194**:855–61.
- 45 Shellenberger TD, Wang M, Gujrati M *et al.* BRAK/CXCL14 is a potent inhibitor of angiogenesis and a chemotactic factor for immature dendritic cells. *Cancer Res* 2004; **64**:8262–70.
- 46 Shurin GV, Ferris R, Tourkova IL *et al.* Loss of new chemokine CXCL14 in tumor tissue is associated with low infiltration by dendritic cells (DC), while restoration of human CXCL14 expression in tumor cells causes attraction of DC both in vitro and in vivo. *J Immunol* 2005; **174**:5490–8.
- 47 Schwarze SR, Luo J, Isaacs WB, Jarrard DF. Modulation of CXCL14 (BRAK) expression in prostate cancer. *Prostate* 2005; **64**:67–74.
- 48 Matloubian M, David A, Engel S, Ryan JE, Cyster JG. A transmembrane CXC chemokine is a ligand for HIV-coreceptor Bonzo. *Nat Immunol* 2000; **1**:298–304.
- 49 Wilbanks A, Zondlo SC, Murphy K *et al.* Expression cloning of the STRL33/BONZO/TYMSTR ligand reveals elements of CC, CXC, and CX3C chemokines. *J Immunol* 2001; **166**:5145–54.
- 50 Shimaoka T, Nakayama T, Fukumoto N *et al.* Cell surface-anchored SR-PSOX/CXC chemokine ligand 16 mediates firm adhesion of CXC chemokine receptor 6-expressing cells. *J Leukoc Biol* 2004; **75**:267–74.
- 51 Gough PJ, Garton KJ, Wille PT, Rychlewski M, Dempsey PJ, Raines EW. A disintegrin and metalloproteinase 10-mediated cleavage and shedding regulates the cell surface expression of CXC chemokine ligand 16. *J Immunol* 2004; **172**:3678–85.
- 52 Abel S, Hundhausen C, Mentlein R *et al.* The transmembrane CXC-chemokine ligand 16 is induced by IFN-gamma and TNF-alpha and shed by the activity of the disintegrin-like metalloproteinase ADAM10. *J Immunol* 2004; **172**:6362–72.
- 53 Blanpain C, Migeotte I, Lee B *et al.* CCR5 binds multiple CC-chemokines: MCP-3 acts as a natural antagonist. *Blood* 1999; **94**:1899–905.
- 54 Ogilvie P, Bardi G, Clark-Lewis I, Baggiolini M, Uguccioni M. Eotaxin is a natural antagonist for CCR2 and an agonist for CCR5. *Blood* 2001; **97**:1920–4.
- 55 Ogilvie P, Paoletti S, Clark-Lewis I, Uguccioni M. Eotaxin-3 is a natural antagonist for CCR2 and exerts a repulsive effect on human monocytes. *Blood* 2003; **102**:789–94.
- 56 Zingoni A, Soto H, Hedrick JA *et al.* The chemokine receptor CCR8 is preferentially expressed in Th2 but not Th1 cells. *J Immunol* 1998; **161**:547–51.
- 57 Nomiya H, Fukuda S, Iio M, Tanase S, Miura R, Yoshie O. Organization of the chemokine gene cluster on human chromosome 17q11.2 containing the genes for CC chemokine MIP1-1, HCC-2, HCC-1, LEC, and RANTES. *J Interferon Cytokine Res* 1999; **19**:227–34.
- 58 Schulz-Knappe P, Magert HJ, Dewald B *et al.* HCC-1, a novel chemokine from human plasma. *J Exp Med* 1996; **183**:295–9.
- 59 Howard OM, Dong HF, Shirakawa AK, Oppenheim JJ. LEC induces chemotaxis and adhesion by interacting with CCR1 and CCR8. *Blood* 2000; **96**:840–5.
- 60 Nomiya H, Hieshima K, Nakayama T *et al.* Human CC chemokine liver-expressed chemokine/CCL16 is a functional ligand for CCR1, CCR2 and CCR5, and constitutively expressed by hepatocytes. *Int Immunol* 2001; **13**:1021–9.
- 61 Pardigol A, Forssmann U, Zucht HD *et al.* HCC-2, a human chemokine: gene structure, expression pattern, and biological activity. *Proc Natl Acad Sci USA* 1998; **95**:6308–13.
- 62 Youn BS, Zhang SM, Broxmeyer HE *et al.* Characterization of CKbeta8 and CKbeta8-1: two alternatively spliced forms of human beta-chemokine, chemoattractants for neutrophils, monocytes, and lymphocytes, and potent agonists at CC chemokine receptor 1. *Blood* 1998; **91**:3118–26.
- 63 Appay V, Rowland-Jones SL. RANTES: a versatile and controversial chemokine. *Trends Immunol* 2001; **22**:83–7.
- 64 Menten P, Wuyts A, Van Damme J. Macrophage inflammatory protein-1. *Cytokine Growth Factor Rev* 2002; **13**:455–81.
- 65 Hirashima M, Ono T, Nakao M *et al.* Nucleotide sequence of the third cytokine LD78 gene and mapping of all three LD78 gene loci to human chromosome 17. *DNA Seq* 1992; **3**:203–12.
- 66 Townson JR, Barcellos LF, Nibbs RJ. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol* 2002; **32**:3016–26.
- 67 Colobran R, Adreani P, Ashhab Y *et al.* Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. *J Immunol* 2005; **174**:5655–64.
- 68 Maurer M, von Stebut E. Macrophage inflammatory protein-1. *Int J Biochem Cell Biol* 2004; **36**:1882–6.
- 69 Tasaki Y, Fukuda S, Iio M *et al.* Chemokine PARC gene (SCYA18) generated by fusion of two MIP-1alpha/LD78alpha-like genes. *Genomics* 1999; **55**:353–7.
- 70 Schutysse E, Richmond A, Van Damme J. Involvement of CC chemokine ligand 18 (CCL18) in normal and pathological processes. *J Leukoc Biol* 2005; **78**:14–26.
- 71 Nibbs RJ, Salcedo TW, Campbell JD *et al.* C-C chemokine receptor 3 antagonism by the beta-chemokine macrophage inflammatory protein 4, a property strongly enhanced by an amino-terminal alanine-methionine swap. *J Immunol* 2000; **164**:1488–97.
- 72 Gunn MD, Tangemann K, Tam C, Cyster JG, Rosen SD, Williams LT. A chemokine expressed in lymphoid high endothelial venules

- promotes the adhesion and chemotaxis of naive T lymphocytes. *Proc Natl Acad Sci USA* 1998; **95**:258–63.
- 73 Kim CH, Pelus LM, White JR, Applebaum E, Johanson K, Broxmeyer HE. CK beta-11/macrophage inflammatory protein-3 beta/EBI1-ligand chemokine is an efficacious chemoattractant for T and B cells. *J Immunol* 1998; **160**:2418–24.
- 74 Homey B, Alenius H, Muller A *et al.* CCL27–CCR10 interactions regulate T cell-mediated skin inflammation. *Nat Med* 2002; **8**:157–65.
- 75 Yamashita U, Kuroda E. Regulation of macrophage-derived chemokine (MDC, CCL22) production. *Crit Rev Immunol* 2002; **22**:105–14.
- 76 Imai T, Baba M, Nishimura M, Kakizaki M, Takagi S, Yoshie O. The T cell-directed CC chemokine TARC is a highly specific biological ligand for CC chemokine receptor 4. *J Biol Chem* 1997; **272**:15036–42.
- 77 Mizoue LS, Bazan JF, Johnson EC, Handel TM. Solution structure and dynamics of the CX3C chemokine domain of fractalkine and its interaction with an N-terminal fragment of CX3CR1. *Biochemistry* 1999; **38**:1402–14.
- 78 Bazan JF, Bacon KB, Hardiman G *et al.* A new class of membrane-bound chemokine with a CX3C motif. *Nature* 1997; **385**:640–4.
- 79 Yang D, Chertov O, Bykovskaia SN *et al.* Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* 1999; **286**:525–8.
- 80 Schutyser E, Struyf S, Van Damme J. The CC chemokine CCL20 and its receptor CCR6. *Cytokine Growth Factor Rev* 2003; **14**:409–26.
- 81 Kunkel EJ, Campbell JJ, Haraldsen G *et al.* Lymphocyte CC chemokine receptor 9 and epithelial thymus-expressed chemokine (TECK) expression distinguish the small intestinal immune compartment: epithelial expression of tissue-specific chemokines as an organizing principle in regional immunity. *J Exp Med* 2000; **192**:761–8.
- 82 Vicari AP, Figueroa DJ, Hedrick JA *et al.* TECK: a novel CC chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development. *Immunity* 1997; **7**:291–301.
- 83 Lazarus NH, Kunkel EJ, Johnston B, Wilson E, Youngman KR, Butcher EC. A common mucosal chemokine (mucosae-associated epithelial chemokine/CCL28) selectively attracts IgA plasmablasts. *J Immunol* 2003; **170**:3799–805.
- 84 Pan J, Kunkel EJ, Gossler U *et al.* A novel chemokine ligand for CCR10 and CCR3 expressed by epithelial cells in mucosal tissues. *J Immunol* 2000; **165**:2943–9.
- 85 Wang W, Soto H, Oldham ER *et al.* Identification of a novel chemokine (CCL28), which binds CCR10 (GPR2). *J Biol Chem* 2000; **275**:22313–23.
- 86 Yoshida T, Ishikawa I, Ono Y, Imai T, Suzuki R, Yoshie O. An activation-responsive element in single C motif-1/lymphotactin promoter is a site of constitutive and inducible DNA–protein interactions involving nuclear factor of activated T cell. *J Immunol* 1999; **163**:3295–303.
- 87 Kelner GS, Kennedy J, Bacon KB *et al.* Lymphotactin: a cytokine that represents a new class of chemokine. *Science* 1994; **266**:1395–9.

## 4.5. CAPÍTOL V

### OBJECTIU:

Ampliar el coneixement de la complexitat de la superfamília de les quimiocines mitjançant la recopilació dels polimorfismes més rellevants dels seus membres i de les formes generades per tall-i-unió alternatiu. Establir la relació d'aquests fenòmens amb patologia.

### ARTICLE:

**The chemokine network. II. On how polymorphisms and alternative splicing increase the number of molecular species and configure intricate patterns of disease susceptibility.**

R. Colobran, R. Pujol-Borrell, M<sup>a</sup> P. Armengol and M. Juan.

*Clinical and Experimental Immunology*, **2007**, 150: 1–12.



#### 4.5.1. CONSIDERACIONS PRÈVIES SOBRE L'ARTICLE

L'elevada variabilitat i complexitat de la parella de quimiocines CCL4-CCL4L descrita en els capítols I i III, causada principalment per polimorfismes i fenòmens de tall-i-unió alternatiu, va fer que analitzéssim la presència d'aquests dos fenòmens a la resta dels membres de la superfamília de les quimiocines.

#### 4.5.2. RESUM DE DADES DEL CINQUÈ CAPÍTOL:

L'objectiu era arribar a tenir una visió general de l'augment de variabilitat en la complexa xarxa de quimiocines humanes causada per polimorfismes genètics i per variants originades per tall-i-unió alternatiu, i de la seva possible repercussió en patologia. Un cop més, la bibliografia sobre aquest aspecte era pràcticament nul·la. Si bé trobàvem múltiples publicacions sobre estudis que analitzaven polimorfismes concrets en patologies concretes, mancaven estudis més generals que ens proporcionessin aquesta visió global. D'aquí la realització d'aquesta segona revisió, que es presenta a continuació. El que hem fet és destriar d'entre els centenars de polimorfismes descrits en les quimiocines (bàsicament SNPs), aquells que tenen una repercussió en patologia. Per altra banda, hem recopilat tota la informació existent sobre variants de quimiocines originades per fenòmens de tall-i-unió alternatiu.

Durant l'estudi també s'analitzen certes dades contradictòries que s'han trobat en les diverses publicacions consultades. A la part final es discuteix sobre la necessitat de futurs estudis que clarifiquin alguns aspectes encara confusos o poc definits sobre els polimorfismes de quimiocines i la susceptibilitat a malaltia. Pel que fa a les variants originades per tall-i-unió alternatiu, encara existeixen molt poques dades sobre la seva repercussió funcional i en malaltia, per la qual cosa caldran més estudis que aprofundeixin sobre aquest aspecte.



# The chemokine network. II. On how polymorphisms and alternative splicing increase the number of molecular species and configure intricate patterns of disease susceptibility

R. Colobran, R. Pujol-Borrell,  
M. P. Armengol and M. Juan  
*Laboratory of Immunobiology for Research and  
Application to Diagnosis (LIRAD), Tissue and  
Blood Bank (BST), Institut d'Investigació en  
Ciències de la Salut Germans, Trias i Pujol  
(IGTP), Badalona, and Department of Cell  
Biology, Physiology and Immunology, Universitat  
Autònoma de Barcelona (UAB), Bellaterra,  
Barcelona, Spain*

Accepted for publication 24 July 2007

Correspondence: Manel Juan Otero, Hospital  
Germans Trias i Pujol, Edifici de Recerca,  
Carretera de Can Ruti, camí de les escoles,  
s/n, 08916 Badalona, Spain.

E-mail: mjuan.liradbst.germanstrias@gencat.net

## Summary

In this second review on chemokines, we focus on the polymorphisms and alternative splicings and on their consequences in disease. Because chemokines are key mediators in the pathogenesis of inflammatory, autoimmune, vascular and neoplastic disorders, a large number of studies attempting to relate particular polymorphisms of chemokines to given diseases have already been conducted, sometimes with contradictory results. Reviewing the published data, it becomes evident that some chemokine genes that are polymorphic have alleles that are found repeatedly, associated with disease of different aetiologies but sharing some aspects of pathogenesis. Among CXC chemokines, single nucleotide polymorphisms (SNPs) in the CXCL8 and CXCL12 genes stand out, as they have alleles associated with many diseases such as asthma and human immunodeficiency virus (HIV), respectively. Of CC chemokines, the stronger associations occur among alleles from SNPs in CCL2 and CCL5 genes and a number of inflammatory conditions. To understand how chemokines contribute to disease it is also necessary to take into account all the isoforms resulting from differential splicing. The first part of this review deals with polymorphisms and the second with the diversity of molecular species derived from each chemokine gene due to alternative splicing phenomena. The number of molecular species and the level of expression of each of them for every chemokine and for each functionally related group of chemokines reaches a complexity that requires new modelling algorithms akin to those proposed in systems biology approaches.

**Keywords:** chemokines, human, polymorphisms, splicing, variability

## Increasing chemokine variability: polymorphisms and alternative splicing

In a first review [1] we examined data which indicated that, during evolution, the variability of the chemokine superfamily grew in complexity, and we took advantage of the conservation of physiological functions among chemokines located in the different genetic clusters and miniclusters to improve our perspective of their functions. As for many gene families, the main mechanism that has generated this diversity of chemokines is gene duplication, which is particularly evident in the chemokine clusters. However, another important mechanism by which variation has been increasing at the genomic level is the existence of single nucleotide polymorphisms (SNPs), which are the most common form of DNA sequence variation. SNPs are highly abundant, stable and distributed throughout the genome. SNPs are an increas-

ingly important tool for the study of the structure and history of human genome and they are also useful polymorphic markers to investigate genetic susceptibility to disease or to pharmacological sensitivity [2]. Other types of polymorphisms such as deletion/insertion polymorphisms (DIPs), copy number polymorphisms (CNPs) or those due to repeated elements (as minisatellites and microsatellites) also contribute importantly to the genomic variation but their distribution is more restricted. In addition to DNA sequence variation, alternative mRNA splicing is becoming recognized increasingly as an important mechanism for the generation of structural and functional variability in proteins. Several studies indicate that alternative splicing in humans is more the rule than the exception: primary transcripts from more than 50% of all human genes undergo alternative splicing, with a bias towards genes that are expressed in the nervous and immune systems [3,4].



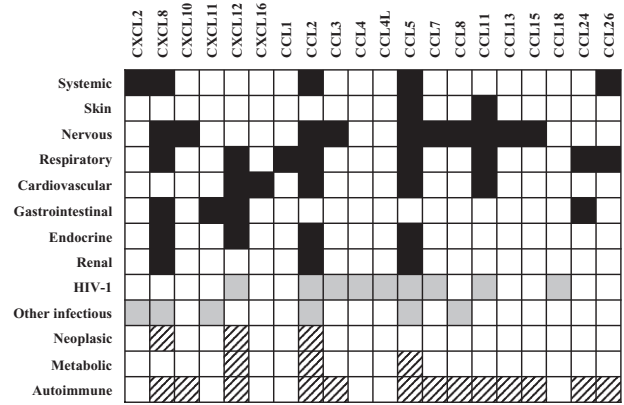
In this second review, we focused upon the polymorphisms and disease associations of chemokine genes as well as variations in splicing which should be taken into account in order to understand these disease associations more clearly. We proceeded by collecting all information available in public databases, organized it by families following the systematic nomenclature [5], and finally we highlighted the cases in which disease association is stronger. In the course of this process we also analysed data available on isoforms generated by differential splicing. Even though we can expect that new data will still be produced on the polymorphisms and isoforms of chemokines, we now have a picture of their complexity and we can begin to discern patterns of disease association; this is also the subject of this review.

### Polymorphisms and disease in the human chemokine superfamily

Polymorphisms in the genes of the immune system can influence the immune response markedly, human leucocyte antigen (HLA) genes being the paradigm. After the HLA genes, chemokine genes are probably one of the most polymorphic sets of genes in the immune system and it is becoming increasingly clear that chemokine polymorphisms influence the immune response to a remarkable extent. As the genome project progressed and the abundance of SNPs became evident, databases began to record SNPs and now millions of them are registered. However, the quality of the initial data contained in these databases had been questioned, because a considerable proportion of the initial SNPs may simply represent sequencing errors. Fortunately, the validation status of SNPs is improving and in this review we have included only well-documented and functionally relevant SNPs. The number of reports on disease-associated SNPs including members of the chemokine superfamily is increasing and will probably continue to rise during the next few years, as the importance of chemokines in the immune response gains recognition. As has already been documented for cytokines, the majority of SNPs found in the chemokines genes or their receptors are not located in the coding sequence but either in the promoter, the introns or the 3' untranslated regions, and they can affect all aspects of gene expression and mRNA levels. Interestingly, most polymorphisms associated with disease in the chemokine superfamily affect their inflammatory members, thus confirming that they are the genes under stronger evolutionary pressure (Fig. 1).

#### Human CXC chemokines

Several interesting polymorphisms affecting both inflammatory and homeostatic CXC ligands have been described (Table 1), CXCL8 [interleukin (IL)-8] and CXCL12 stromal cell-derived factor (SDF-1) being the chemokines that accumulate most of them.



**Fig. 1.** Overview of chemokine polymorphisms and disease. Columns represent chemokines with disease-involved polymorphisms. Rows represent different disease categories: upper lines refer to diseases grouped by systems (black squares show polymorphism involvement), middle lines refer to infectious diseases (grey squares) and lower lines refer to other interesting physiopathological groups of diseases (dashed squares).

CXCL8, a proinflammatory chemokine, is a potent chemoattractant for neutrophils, basophils and T lymphocytes. High levels of CXCL8 have been detected in biofluids from various acute inflammatory diseases, which is in keeping with neutrophilic infiltration into inflammatory sites as one of the hallmarks of acute inflammation. Among all CXCL8 described SNPs, the presence of  $-251A/T$  in the transcription start site is known to exert a strong influence on protein synthesis. The distribution of this SNP shows a remarkable heterogeneity among world populations [6] and has been associated with a spectrum of diseases (for references see Table 1): (a) airway diseases such as asthma, respiratory syncytial virus (RSV) infection and oral squamous cell carcinoma. Anecdotally, the inflammatory prone allele A may influence the initiation or characteristics of the smoking habit [7]; (b) gastrointestinal diseases such as *Clostridium difficile* and enteroaggregative *Escherichia coli* diarrhoea, *Helicobacter pylori*-induced gastric ulcer, atrophic gastritis, severe acute pancreatitis and gastrointestinal tract cancer; (c) central nervous system (CNS) diseases such as Parkinson's, multiple sclerosis (MS) and multiple system atrophy; and (d) a miscellany of diseases such as AIDS-related Kaposi's sarcoma and acute pyelonephritis has also shown to be influenced by the  $-251A/T$  CXCL8 polymorphism. However, in spite of the numerous studies on this polymorphism, data are still far from clear. Since the publication of the original association of the CXCL8  $-251A/T$  polymorphism [8,9], there have been reports showing higher CXCL8 production by the A allele [8,10,11] while others showed higher production by allele T [12]. These contradictory data have their counterpart in association studies: the  $-251 T$  allele frequency has been increased significantly in asthma [13] and reduced significantly in RSV bronchiolitis [8]. In fact, it

**Table 1.** Polymorphisms and disease in the human CXC subfamily.

Ligand	Polymorphism	Location	Symbol	Disease involved	References	
CXCL2	Tandem repeat -665(AC) <sub>n</sub>	Promoter		Severe sepsis	[28]	
CXCL8	Microsatellite		D4S2641	Diffuse parabronchiolitis	[64]	
		SNP -845 (C/T)	Promoter	rs2227532	SLE nephritis	[16]
		SNP -251 (A/T)	Promoter	rs4073	Asthma	[13]
				RSV infection	[8, 9, 14]	
				Smoking behaviour	[7]	
				Oral squamous cell carcinoma	[65]	
				EAEC diarrhoea	[10]	
				<i>Clostridium difficile</i> diarrhoea	[66]	
				<i>Helicobacter pylori</i> -induced gastric diseases	[11, 67–70]	
				Acute pancreatitis severity	[71]	
				Gastric cancer	[12, 72, 73]	
				Colorectal cancer	[74]	
				Prostate cancer	[75]	
				Parkinson's disease	[76]	
				Multiple sclerosis	[77]	
			Multiple system atrophy	[78]		
			AIDS-related Kaposi's sarcoma	[79]		
			Acute pyelonephritis	[17]		
	SNP +781 (C/T)	Intron 1	rs2227306	Asthma	[13, 15]	
	SNP +1633 (C/T)	Intron 3	rs2227543	Asthma	[13]	
	SNP +2767 (A/T)	3'UTR	rs1126647	Asthma	[13]	
				Nephritis in cutaneous vasculitis	[18]	
				Acute pyelonephritis	[17]	
				Behcet's disease	[80]	
	Haplotypes					
CXCL10	Haplotypes			Multiple sclerosis	[81]	
CXCL11	DIP -599del5	Promoter		Hepatitis C virus (HCV) infection	[29]	
CXCL12	SNP +801 (G/A)	3'UTR	rs1801157	HIV-1 infection	[20, 22–24]	
				Atherosclerosis in HIV patients	[82]	
				Breast and lung cancer	[83–85]	
				Acute myeloid leukemia	[86]	
				Lymphoma	[87]	
				Chronic myeloproliferative disease	[88]	
				Liver transplantation	[89]	
				Type 1 diabetes	[90]	
CXCL16	SNP +599 (C/T)	Exon 4	rs2277680	Coronary artery stenosis	[30]	

seems that the differences in CXCL8 expression are not linked directly to the -251A/T polymorphism. In one haplotype-based association study, Hacking *et al.* [14] have shown that there are two main CXCL8 haplotypes including six SNPs (-251A/T, +396G/T, +781T/C, +1238delA/insA, +1633T/C, +2767T/A), constituting the so-called haplotype 2 (A/G/T/delA/T/T), the haplotype associated with significantly higher CXCL8 transcription levels relative to the mirror haplotype 1. Strikingly, the -251A allele present on the high producer haplotype had no significant effect on the allele-specific level of transcription when analysed in reporter gene experiments. This indicates that the functional allele might be in linkage disequilibrium (LD) with haplotype 2 and that the -251A/T is not the functional SNP. Four SNPs of the previously described haplotypes (-251A/T, +781T/C, +1633T/C and +2767T/A) have been found to be

associated with asthma in different studies [13,15]. These multiple SNPs associations are due probably to the existence of a very tight LD among them. The -845C/T SNP in the CXCL8 promoter region has been associated with severe systemic lupus erythematosus (SLE) nephritis [16]. Another distant SNP in 3'UTR (+2767A/T) has been associated with acute pyelonephritis [17] and nephritis in cutaneous vasculitis [18].

CXCL12 is a homeostatic CXC chemokine widely expressed which possesses a broad range of actions (from attraction of mature T and B cells to migration of haematopoietic progenitor cells from the bone marrow). CXCL12 plays an especially important role in two non-related diseases such as human immunodeficiency virus (HIV) and cancer, because its receptor (CXCR4) is also the co-receptor used by HIV T-tropic strains and because it is the most

widely expressed chemokine receptor in many different types of cancers [19]. The +801G/A SNP, located in 3'UTR, is the best-studied polymorphism in CXCL12 gene. It has been associated extensively with clinical features of HIV infection but, as in CXCL8-251A/T polymorphism, there are some contradictory reports. The published effect of the mutated allele (-801A) ranges from strong protection of HIV infection progression to AIDS [20,21] to enhanced progression to AIDS and shorter survival [22,23]. Even though it was proposed originally that the -801A allele was associated with higher CXCL12 production [20], later studies indicated the opposite [24], and other reports claimed that there were no differences in the CXCL12 production by the A or G alleles [25,26]. A recent haplotype-based study [27] demonstrated that other polymorphisms in LD with the CXCL12 + 801G/A SNP, rather than CXCL12 + 801G/A itself, are responsible for the different transcription levels. Therefore, the discrepancy among the previous epidemiological studies may be attributed to the haplotype structures and frequencies in the studied populations. CXCL12 + 801G/A has also been associated with many different types of cancers such as breast and lung cancer, acute myeloid leukaemia, lymphoma and chronic myeloproliferative disease.

Relatively few reports deal with the effect of polymorphisms of other CXC chemokines on disease. A short tandem repeat (STR) in CXCL2 may contribute to the development of severe sepsis [28], one haplotype in CXCL10 possibly contributes to reduce the rate of progression in MS patients, a 5-base pairs (bp) deletion in the promoter of CXCL11 may favour hepatitis C virus (HCV) infection to evolve towards chronicity [29] and, finally, a SNP in exon 4 of CXCL16, leading to an amino acid change (V200A), seemed to influence the severity of coronary artery stenosis [30].

### Human CC chemokines

As many as 14 of the 26 members of CC chemokine subfamily have polymorphisms associated with disease, and most affect the inflammatory chemokines (Table 2). CCL2 [monocyte chemoattractant protein (MCP)-1] attracts specifically monocytes and memory T cells and tissue expression is found in a large variety of diseases characterized by mononuclear cell infiltration, with an essential role in atherosclerosis and multiple sclerosis. CCL2 has a SNP located in the 5' distal regulatory region (-2518G/A) and it seems clear that the -2518G allele is associated with an increased CCL2 production (at both mRNA and protein levels) [31-34]. This -2518G/A polymorphism has been associated with a large variety of diseases: (a) systemic inflammatory diseases such as systemic lupus erythematosus (SLE), juvenile rheumatoid arthritis, systemic sclerosis and HLA-B27-associated acute uveitis; (b) conditions affecting the kidney such as renal transplantation, long-term haemodialysis and IgA nephropathy; (c) heart diseases such as myocardial

infarction, coronary artery disease and the cardiomyopathy of Chagas' disease; (d) CNS diseases such as Alzheimer's and major depression; (e) endocrine diseases such as type 1 and type 2 diabetes; (f) infectious diseases such as those caused by HIV-1, HCV, HBV and *Mycobacterium tuberculosis*; and (g) other diseases such as breast cancer and asthma. It is important to note that there are also many negative reports showing a lack of association of this SNP with various diseases (including some of those cited previously as associated diseases).

CCL3 [macrophage inflammatory protein (MIP)-1 $\alpha$ ], CCL4 (MIP-1 $\beta$ ), CCL4L [lymphocyte activation gene (LAG)-1] and CCL5 [regulated upon activation normal T cell expressed and secreted (RANTES)] have a diversity of polymorphisms that have an important impact on susceptibility to HIV-1 infection. This is not surprising, as they are ligands of the CCR5 receptor, which is the co-receptor used by HIV M-tropic strains to enter into the cells. Haplotypes defined on the region containing the genes CCL18, CCL3 and CCL4 (chromosome 17 q11-q21) have been found to be associated with HIV infection susceptibility and progression [35]. Although CCL18 has not yet been implicated in HIV-1/AIDS pathogenesis and its receptor is not known, this genetic analysis points to this gene as a candidate for modulating HIV-1 pathogenesis. CCL5 haplotypes have also been shown to influence the clinical progression of HIV infection [36,37]. Two interesting SNPs in the CCL4L gene have been associated with different aspects of HIV-1 infection: (a) the +590A/G is located at the intron 2 acceptor splice site. The G allele disrupts the original acceptor splice site and provokes a new complex transcription pattern. This allele modifies susceptibility to HIV-1 infection [38]. (b) The +59C/T is located in exon 2, leading to an amino acid change (R22H). The H variant has been associated with a lower overall survival of HIV-1 infected individuals [39]. Three CCL5 individual SNPs have also been associated with HIV-1 infection, two of them located in the promoter region (-403G/A and -28C/G) and the other in intron 1 (In1.1T/C). It has been demonstrated clearly that the -403A and -28G alleles enhance CCL5 production [40-43] and, conversely, the In1.1C allele reduces CCL5 gene transcription [37]. The two CCL5 promoter polymorphisms, -403G/A and -28C/G, have also been associated with a variety of other diseases such as allergic diseases (i.e. asthma, atopy, allergic rhinitis and atopic dermatitis), inflammatory diseases [i.e. SLE, MS, rheumatoid arthritis (RA), sarcoidosis and polymyalgia rheumatica] and infectious diseases (i.e. HIV-1 and HCV). Additionally, the -403G/A polymorphism has been found to be associated with metabolic risk-related conditions such as hypercholesterolaemia, coronary arteriosclerosis and cardiac mortality in type 2 diabetes.

CCL11 (eotaxin-1), CCL24 (eotaxin-2) and CCL26 (eotaxin-3) are CCR3 ligands and potent eosinophil chemoattractants, playing a fundamental role in asthma and other allergic diseases and in eosinophil-associated

**Table 2.** Polymorphisms and disease in the human CC subfamily.

Ligand	Polymorphism	Location	Symbol	Disease involved	References
CCL1	SNP (A/T)	Intron 2	rs2282691	Chronic obstructive pulmonary disease	[91]
CCL2	SNP -2518 (G/A)	Promoter	rs1024611	Systemic sclerosis	[92]
				Asthma	[93, 94]
				Systemic lupus erythematosus	[95–97]
				Juvenile rheumatoid arthritis	[98]
				Renal transplantation	[99]
				Breast cancer	[100]
				Long-term haemodialysis	[101]
				IgA nephropathy	[102]
				HLA-B27 associated disease	[103]
				Coronary artery disease	[104, 105]
				Myocardial infarction	[106]
				Alzheimer's disease association	[107]
				Major depressive disorder	[108]
				Type 1 diabetes	[109]
				Type 2 diabetes	[110]
				HIV-1 infection	[33]
				Pulmonary tuberculosis	[111]
				Cardiomyopathy in human Chagas' disease	[112]
				Hepatitis B virus (HBV) clearance	[113]
				Hepatitis C virus (HCV) severity	[34]
	Haplotypes			Multiple sclerosis	[114, 115]
	Haplotypes			HIV-1 infection	[116]
CCL3	Haplotypes			Multiple sclerosis	[114, 115, 117]
	Haplotypes			HIV-1 infection	[35, 36]
CCL4	Haplotypes			HIV-1 infection	[35]
CCL4L	SNP +59 (C/T)	Exon 2	rs3744595	HIV-1 infection	[39]
	SNP +590 (A/G)	Intron 2	rs4796195	HIV-1 infection	[38]
CCL5	SNP -403 (G/A)	Promoter	rs2107538	Allergic rhinitis	[118]
				Atopy and asthma	[119–121]
				Atopic dermatitis	[40–42]
				Renal damage in SLE	[122]
				Rheumatoid arthritis	[123, 124]
				Multiple sclerosis	[125]
				HIV-1 infection	[36, 126, 127]
				HCV infection	[128, 129]
				Sarcoidosis	[130]
				Coronary arteriosclerosis	[131]
				Hypercholesterolaemia	[132]
				Cardiac mortality in type 2 diabetes	[133]
	SNP -28 (C/G)	Promoter	rs2280788	Allergic rhinitis	[118]
				Asthma	[134, 135]
				Nephropathy in type 2 diabetes	[136]
				HIV-1 infection	[36, 43, 127]
				Multiple sclerosis	[125]
				Systemic lupus erythematosus	[137]
				Atopic dermatitis	[42]
	SNP In1.1 (T/C)	Intron 1	rs2280789	Cardiac mortality in type 2 diabetes	[133]
				HIV-1 infection	[37, 127, 138]
				HIV-1 infection	[36]
				Type 1 diabetes	[139]
CCL7	Microsatellite	Promoter		Multiple sclerosis	[140, 141]
	Haplotypes			HIV-1 infection	[116]
CCL8	SNP +11 (A/C)	Exon 3	rs1133763	HCV infection	[128]
	Haplotypes			Multiple sclerosis	[114]

**Table 2.** *Continued*

Ligand	Polymorphism	Location	Symbol	Disease involved	References
CCL11	SNP -576 (C/T)	Promoter	rs4795896	Asthma	[44, 45]
	SNP -426 (C/T)	Promoter	rs16969415	Atopic dermatitis	[45, 142]
				Asthma	[45]
	SNP -384 (A/G)	Promoter	rs17809012	Atopic dermatitis	[44, 142]
				Asthma	[45]
	SNP +67 (G/A)	Exon 1	rs3744508	Asthma	[46, 47]
				Myocardial infarction	[143]
	Haplotypes		Multiple sclerosis	[114]	
	Haplotypes		HIV-1 infection	[116]	
CCL13	Haplotypes			Multiple sclerosis	[114]
CCL15	Haplotypes			Multiple sclerosis	[114, 115]
CCL18	Haplotypes			HIV-1 infection	[35]
CCL24	SNP +179 (T/C)	Intron 1	rs2302004	Asthma	[144]
				Ulcerative colitis	[145]
	SNP +275 (C/T)	Intron 1	rs2302005	Asthma	[144]
				Ulcerative colitis	[145]
	SNP +1265 (A/G)	Intron 2	rs11465310	Asthma	[47, 48]
CCL26	SNP +77 (C/T)	Intron 2	rs2240478	Asthma	[144]
	SNP +1577 (G/A)	Intron 3	rs6965556	Rheumatoid arthritis	[146]
	SNP +2497 (T/G)	3'UTR	rs2302009	Rheumatoid arthritis	[146]
				Asthma	[144]
			Allergic rhinitis	[147]	

gastrointestinal diseases. Not unexpectedly, SNPs in these three chemokines have been found to be associated with allergic diseases such as asthma, allergic rhinitis and atopic dermatitis. Four SNPs of the CCL11 gene have been found to be associated independently with asthma in several studies [44–47]. Three of them are located in the promoter region (–576C/T, –426C/T and –384 A/G) and the other (+67G/A) in the signal peptide (exon 1) leading to an amino acid change (T23A). Interestingly, both the –384G and +67A alleles are associated with lower CCL11 production [44,46]. The three polymorphisms of the CCL24 gene associated with asthma are intronic SNPs: +179TC and +275C/T in intron 1 and +1265A/G in intron 2. There are data indicating that the +1265A allele is associated with lower CCL24 levels than the G allele [48]). Finally, CCL26 has two SNPs affecting asthma differently: the +2497T/G (in the 3'UTR region) have been associated with susceptibility and the +77C/T (in intron 2) seem to play a critical role in attracting eosinophils and maintaining high IgE levels.

Regarding the C and CX3C subfamilies, no relevant polymorphisms in their members have so far been described.

### Transcriptional variability: alternative splicing in the chemokine superfamily

The mRNA of several chemokines is known to undergo alternative splicing (Table 3), some of them with repercussions in the molecular activity and/or in the tissue distribution of the differentially spliced variants. However, to date,

there are no reports on their implication in disease pathogenesis.

CXCL12 is the only CXC chemokine known to generate isoforms by alternative splicing. The two main splice forms of CXCL12 (SDF-1 $\alpha$  and SDF-1 $\beta$ ) have similar amino acid sequences except for the presence of four additional amino acids at the carboxy terminus of SDF-1 $\beta$ . Both isoforms display a similar tissue expression pattern, but SDF-1 $\alpha$  mRNA can be detected in the adult human brain, whereas SDF-1 $\beta$  cannot. The two isoforms are subjected to different proteolytic processing, and this fact could explain functional differences [49]. Recently, four additional human SDF-1 isoforms derived from alternative splicing events have been identified (SDF-1 $\gamma$ , SDF-1 $\delta$ , SDF-1 $\epsilon$  and SDF-1 $\phi$ ), showing some differential distribution of tissue expression [50].

CCL4 and CCL4L, two closely related chemokines, have different isoforms due to alternative splicing. Both chemokines have exon 2 skipped variants that keep only the two first amino acids from the original protein due to a frameshift in the new junction between exon 1 and exon 3. Additionally, CCL4L2 (an allelic variant of CCL4L) has a nucleotide change in the acceptor splice site of intron 2 leading to a complex transcription pattern due to multiple usage of new alternative acceptor splice sites surrounding the original mutated one [38].

Two alternative splice isoforms of CCL20 have been identified, resulting from the alternative usage of two potential acceptor splice sites separated by three nucleotides in the junction of intron 1 and exon 2. The longer form

**Table 3.** Alternative splicing in the human chemokine superfamily.

Ligand	Splicing phenomena	Isoforms	Length	Isoforms identity	Refs
CXCL12	Exon 4 alternative splicing	SDF-1 $\alpha$	68 aa	All variants share the same first three exons but contain different fourth exons	[49, 50, 148]
		SDF-1 $\beta$	72 aa		
		SDF-1 $\gamma$	98 aa		
		SDF-1 $\delta$	119 aa		
		SDF-1 $\epsilon$	69 aa		
		SDF-1 $\phi$	79 aa		
CCL4	Exon 2 skipping	CCL4	69 aa	The $\Delta 2$ isoform keeps only the two first amino acids due to a frameshift	[38]
		CCL4 $\Delta 2$			
CCL4L1	Exon 2 skipping	CCL4L1	69 aa	The $\Delta 2$ isoform keeps only the two first amino acids due to a frameshift	[38]
		CCL4L1 $\Delta 2$	29 aa		
CCL4L2	Alternative acceptor splice sites in exon 3 Exon 2 skipping	CCL4L2	64 aa	The $\Delta 2$ isoforms keep only the two first amino acids due to a frameshift	[38]
		CCL4L2 $\Delta 2$	24 aa	The rest of the isoforms share the same first two exons but contain different third exons	
		CCL4L2b	41 aa		
		CCL4L2b $\Delta 2$	45 aa		
		CCL4L2c	80 aa		
		CCL4L2d	73 aa		
CCL20	Alternative acceptor splice site in exon 2	CCL4L2e	63 aa		
		CCL4L2f	80 aa		
		CCL20 Ala	70 aa	100% (CCL20 Ala has 1 additional aa in N-terminus)	[51, 52]
		CCL20 Ser	69 aa		
CCL23	Alternative acceptor splice site in exon 3	CK $\beta$ 8-1	116 aa	99% (CK8 lacks 17 aa before the two first cysteines)	[53, 54]
		CK $\beta$ 8	99 aa		
CCL27	Alternative first exon usage Intron retention	CCL27	95 aa	PESKY and canonical CCL27 differ only in the first of three exons	[56, 57, 149]
		PESKY	127 aa	The partially spliced and unspliced variants of CCL27 retains the intron 1 and intron 1 and 2, respectively	
		CCL27 unspliced	32 aa		
		CCL27 partially spliced	32 aa		

(CCL20Ala) has an alanine (Ala27) as its predicted N-terminal amino acid, whereas the deletion of Ala27 leads to Ser27 as the predicted N-terminal amino acid in the short form (CCL20Ser) [51,52]. The biological activity of CCL20Ala and CCL20Ser and the tissue-specific preference of different acceptor splice-sites usages are not yet known.

CCL23 has two variants originated by alternative splicing in exon 3: the originally described CK $\beta$ 8 and the splicing variant CK $\beta$ 8-1, which is 17 amino acids longer. The mature proteins CK $\beta$ 8 and CK $\beta$ 8-1 consist of 99 and 116 amino acids, respectively. It has been shown that CK $\beta$ 8 differed from CK $\beta$ 8-1 in the monocyte chemoattraction and in the binding to human formyl peptide-receptor-like-1 (FPRL-1), suggesting that these two CCL23 isoforms could possibly have different a kinetic and specificity of chemotactic function *in vivo* [53,54].

Finally, CCL27 is produced as two splice variants. One of these variants encodes a classical chemokine with an associated signal peptide (CCL27), while the other variant (PESKY) maintains the sequence of the mature chemokine, but the signal peptide has been replaced by an alternative stretch of amino acids that directs this isoform to the nucleus where it modulates transcription. Surprisingly, secreted CCL27 can also reach the nucleus after CCR10-mediated internalization, and in this way directly modulates transcription and influences several cellular processes [55]. Expression studies have revealed differential tissue expression of CCL27 and PESKY. Interestingly, while CCL27 is highly expressed in the placenta, PESKY is expressed mainly in the testes and brain and weakly in the developing embryo [56]. Recently, several novel CCL27 variants have been identified in mouse but their presence in humans has not yet been demonstrated [57].

### Concluding remarks

The high variability of the chemokine superfamily includes mechanisms of genomic and transcriptional variation. There is already a good number of well-described polymorphisms of chemokines with functional relevance and we made a detailed review of those involved significantly in disease. In spite of the many reports on the association of these polymorphisms to diseases, there are still confusing and contradictory data. Many factors in the epidemiological investigation could explain this phenomenon (covered widely in several reviews [58–60]), but it is clear that further studies are necessary to define more clearly the role of genetic variants of chemokines in disease. The recently developed high-throughput methods for SNP genotyping should make it easy to carry out larger association studies using a high number of SNPs, covering from one or a few genes (candidate gene approach) to the whole genome (genome-wide approach). In fact, the single SNP association studies are currently being replaced by the haplotype-based studies using tagSNPs, as this approach ensures the capture

of most of the genetic variation in a relatively transferable manner among global populations [61]. With regard to the alternative splicing phenomena in the chemokine superfamily, several members with different splice variants have been identified but there are still few available data about its functional role. Molecular analyses during the last decade demonstrate that alternative splicing determines the binding properties, intracellular localization, enzymatic activity, protein stability and post-translational modifications of a large number of proteins [62,63]. Efforts are now being directed at establishing the full repertoire of functionally relevant transcript variants generated by alternative splicing, the specific roles of such variants in normal and disease physiology, and how alternative splicing is co-ordinated on a global level to achieve cell- and tissue-specific functions. Although the interaction between all these factors will probably provide us with the true key to understanding their real effect on pathology, future studies will be necessary to achieve all these goals in the chemokine superfamily.

### Acknowledgements

This work was supported by grants from the FIPSE (Fundación para la Investigación y la Prevención del Sida en España) (project 36487/05), FIS (Fondo de Investigaciones Sanitarias) (project PI 02/0104) and PEI (Pla Estratègic d'Investigació) from BST (Banc de Sang i Teixits).

### References

- Colobran R, Pujol-Borrell R, Armengol MP, Juan M. The chemokine network. I. How the genomic organization of chemokines contains clues for deciphering their functional complexity. *Clin Exp Immunol* 2007; **148**:208–17.
- Brookes AJ. The essence of SNPs. *Gene* 1999; **234**:177–86.
- Modrek B, Resch A, Grasso C, Lee C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucl Acids Res* 2001; **29**:2850–9.
- Johnson JM, Castle J, Garrett-Engel P *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 2003; **302**:2141–4.
- Zlotnik A, Yoshie O. Chemokines: a new classification system and their role in immunity. *Immunity* 2000; **12**:121–7.
- Fujihara J, Shiwaku K, Yasuda T *et al.* Variation of interleukin 8 –251 A>T polymorphism in worldwide populations and intra-ethnic differences in Japanese populations. *Clin Chim Acta* 2007; **377**:79–82.
- Ito H, Matsuo K, Hamajima N, Okuma K, Saito T, Tajima K. Significant association of interleukin 8 –251T/A polymorphism with smoking behavior in a Japanese population. *J Hum Genet* 2005; **50**:567–73.
- Hull J, Thomson A, Kwiatkowski D. Association of respiratory syncytial virus bronchiolitis with the interleukin 8 gene region in UK families. *Thorax* 2000; **55**:1023–7.
- Hull J, Ackerman H, Isles K *et al.* Unusual haplotypic structure of IL8, a susceptibility locus for a common respiratory virus. *Am J Hum Genet* 2001; **69**:413–9.

- 10 Jiang ZD, Okhuysen PC, Guo DC *et al.* Genetic susceptibility to enteroaggregative *Escherichia coli* diarrhea: polymorphism in the interleukin-8 promoter region. *J Infect Dis* 2003; **188**:506–11.
- 11 Taguchi A, Ohmiya N, Shirai K *et al.* Interleukin-8 promoter polymorphism increases the risk of atrophic gastritis and gastric cancer in Japan. *Cancer Epidemiol Biomarkers Prev* 2005; **14**:2487–93.
- 12 Lee WP, Tai DI, Lan KH *et al.* The -251T allele of the interleukin-8 promoter is associated with increased risk of gastric carcinoma featuring diffuse-type histopathology in Chinese population. *Clin Cancer Res* 2005; **11**:6431–41.
- 13 Heinzmann A, Ahlert I, Kurz T, Berner R, Deichmann KA. Association study suggests opposite effects of polymorphisms within IL8 on bronchial asthma and respiratory syncytial virus bronchiolitis. *J Allergy Clin Immunol* 2004; **114**:671–6.
- 14 Hacking D, Knight JC, Rockett K *et al.* Increased in vivo transcription of an IL-8 haplotype associated with respiratory syncytial virus disease-susceptibility. *Genes Immun* 2004; **5**:274–82.
- 15 Puthothu B, Krueger M, Heinze J, Forster J, Heinzmann A. Impact of IL8 and IL8-Receptor alpha polymorphisms on the genetics of bronchial asthma and severe RSV infections. *Clin Mol Allergy* 2006; **4**:2.
- 16 Rovin BH, Lu L, Zhang X. A novel interleukin-8 polymorphism is associated with severe systemic lupus erythematosus nephritis. *Kidney Int* 2002; **62**:261–5.
- 17 Artifoni L, Negrisolo S, Montini G *et al.* Interleukin-8 and CXCR1 receptor functional polymorphisms and susceptibility to acute pyelonephritis. *J Urol* 2007; **177**:1102–6.
- 18 Amoli MM, Thomson W, Hajeer AH *et al.* Interleukin 8 gene polymorphism is associated with increased risk of nephritis in cutaneous vasculitis. *J Rheumatol* 2002; **29**:2367–70.
- 19 Zlotnik A. Chemokines and cancer. *Int J Cancer* 2006; **119**:2026–9.
- 20 Winkler C, Modi W, Smith MW *et al.* Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). *Science* 1998; **279**:389–93.
- 21 Vissoci Reiche EM, Ehara Watanabe MA, Bonametti AM *et al.* The effect of stromal cell-derived factor 1 (SDF1/CXCL12) genetic polymorphism on HIV-1 disease progression. *Int J Mol Med* 2006; **18**:785–93.
- 22 van Rij RP, Broersen S, Goudsmit J, Coutinho RA, Schuitemaker H. The role of a stromal cell-derived factor-1 chemokine gene variant in the clinical course of HIV-1 infection. *Aids* 1998; **12**:F85–90.
- 23 Brambilla A, Villa C, Rizzardi G *et al.* Shorter survival of SDF1-3'A/3'A homozygotes linked to CD4+ T cell decrease in advanced human immunodeficiency virus type 1 infection. *J Infect Dis* 2000; **182**:311–5.
- 24 Soriano A, Martinez C, Garcia F *et al.* Plasma stromal cell-derived factor (SDF)-1 levels, SDF1-3'A genotype, and expression of CXCR4 on T lymphocytes: their impact on resistance to human immunodeficiency virus type 1 infection and its progression. *J Infect Dis* 2002; **186**:922–31.
- 25 Arya SK, Ginsberg CC, Davis-Warren A, D'Costa J. In vitro phenotype of SDF1 gene mutant that delays the onset of human immunodeficiency virus disease in vivo. *J Hum Virol* 1999; **2**:133–8.
- 26 Kimura R, Nishioka T, Ishida T. The SDF1-G801A polymorphism is not associated with SDF1 gene expression in Epstein-Barr virus-transformed lymphoblastoid cells. *Genes Immun* 2003; **4**:356–61.
- 27 Kimura R, Nishioka T, Soemantri A, Ishida T. Allele-specific transcript quantification detects haplotypic variation in the levels of the SDF-1 transcripts. *Hum Mol Genet* 2005; **14**:1579–85.
- 28 Flores C, Maca-Meyer N, Perez-Mendez L *et al.* A CXCL2 tandem repeat promoter polymorphism is associated with susceptibility to severe sepsis in the Spanish population. *Genes Immun* 2006; **7**:141–9.
- 29 Helbig KJ, George J, Beard MR. A novel I-TAC promoter polymorphic variant is functional in the presence of replicating HCV *in vitro*. *J Clin Virol* 2005; **32**:137–43.
- 30 Lundberg GA, Kellin A, Samnegard A *et al.* Severity of coronary artery stenosis is associated with a polymorphism in the CXCL16/SR-PSOX gene. *J Intern Med* 2005; **257**:415–22.
- 31 Rovin BH, Lu L, Saxena R. A novel polymorphism in the MCP-1 gene regulatory region that influences MCP-1 expression. *Biochem Biophys Res Commun* 1999; **259**:344–8.
- 32 Jibiki T, Terai M, Shima M *et al.* Monocyte chemoattractant protein 1 gene regulatory region polymorphism and serum levels of monocyte chemoattractant protein 1 in Japanese patients with Kawasaki disease. *Arthritis Rheum* 2001; **44**:2211–2.
- 33 Gonzalez E, Rovin BH, Sen L *et al.* HIV-1 infection and AIDS dementia are influenced by a mutant MCP-1 allele linked to increased monocyte infiltration of tissues and MCP-1 levels. *Proc Natl Acad Sci USA* 2002; **99**:13795–800.
- 34 Muhlbauer M, Bosserhoff AK, Hartmann A *et al.* A novel MCP-1 gene polymorphism is associated with hepatic MCP-1 expression and severity of HCV-related liver disease. *Gastroenterology* 2003; **125**:1085–93.
- 35 Modi WS, Lautenberger J, An P *et al.* Genetic variation in the CCL18–CCL3–CCL4 chemokine gene cluster influences HIV Type 1 transmission and AIDS disease progression. *Am J Hum Genet* 2006; **79**:120–8.
- 36 Gonzalez E, Dhanda R, Bamshad M *et al.* Global survey of genetic variation in CCR5, RANTES, and MIP-1alpha: impact on the epidemiology of the HIV-1 pandemic. *Proc Natl Acad Sci USA* 2001; **98**:5199–204.
- 37 An P, Nelson GW, Wang L *et al.* Modulating influence on HIV/AIDS by interacting RANTES gene variants. *Proc Natl Acad Sci USA* 2002; **99**:10002–7.
- 38 Colobran R, Adreani P, Ashhab Y *et al.* Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. *J Immunol* 2005; **174**:5655–64.
- 39 Capoulade-Metay C, Meyer L, Tran T *et al.* Influence of the R22H variant of macrophage inflammatory protein 1beta/Lag-1 in HIV-1 survival. *Aids* 2005; **19**:831–3.
- 40 Nickel RG, Casolaro V, Wahn U *et al.* Atopic dermatitis is associated with a functional mutation in the promoter of the C-C chemokine RANTES. *J Immunol* 2000; **164**:1612–6.
- 41 Bai B, Tanaka K, Tazawa T, Yamamoto N, Sugiura H. Association between RANTES promoter polymorphism -401A and enhanced RANTES production in atopic dermatitis patients. *J Dermatol Sci* 2005; **39**:189–91.
- 42 Tanaka K, Roberts MH, Yamamoto N, Sugiura H, Uehara M, Hopkin JM. Upregulating promoter polymorphisms of RANTES relate to atopic dermatitis. *Int J Immunogenet* 2006; **33**:423–8.
- 43 Liu H, Chao D, Nakayama EE *et al.* Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. *Proc Natl Acad Sci U S A* 1999; **96**:4581–5.
- 44 Chang HS, Kim JS, Lee JH *et al.* A single nucleotide polymorphism



- on the promoter of eotaxin 1 associates with its mRNA expression and asthma phenotypes. *J Immunol* 2005; **174**:1525–31.
- 45 Raby BA, Van Steen K, Lazarus R, Celedon JC, Silverman EK, Weiss ST. Eotaxin polymorphisms and serum total IgE levels in children with asthma. *J Allergy Clin Immunol* 2006; **117**:298–305.
  - 46 Nakamura H, Luster AD, Nakamura T *et al.* Variant eotaxin: its effects on the asthma phenotype. *J Allergy Clin Immunol* 2001; **108**:946–53.
  - 47 Shin HD, Kim LH, Park BL *et al.* Association of eotaxin gene family with asthma and serum total IgE. *Hum Mol Genet* 2003; **12**:1279–85.
  - 48 Min JW, Lee JH, Park CS *et al.* Association of eotaxin-2 gene polymorphisms with plasma eotaxin-2 concentration. *J Hum Genet* 2005; **50**:118–23.
  - 49 De La Luz Sierra M, Yang F, Narazaki M *et al.* Differential processing of stromal-derived factor-1 $\alpha$  and stromal-derived factor-1 $\beta$  explains functional diversity. *Blood* 2004; **103**:2452–9.
  - 50 Yu L, Cecil J, Peng SB *et al.* Identification and expression of novel isoforms of human stromal cell-derived factor 1. *Gene* 2006; **374**:174–9.
  - 51 Nelson RT, Boyd J, Gladue RP *et al.* Genomic organization of the CC chemokine mip-3 $\alpha$ /CCL20/larc/exodus/SCYA20, showing gene structure, splice variants, and chromosome localization. *Genomics* 2001; **73**:28–37.
  - 52 Tanaka Y, Imai T, Baba M *et al.* Selective expression of liver and activation-regulated chemokine (LARC) in intestinal epithelium in mice and humans. *Eur J Immunol* 1999; **29**:633–42.
  - 53 Youn BS, Zhang SM, Broxmeyer HE *et al.* Characterization of CKbeta8 and CKbeta8-1: two alternatively spliced forms of human beta-chemokine, chemoattractants for neutrophils, monocytes, and lymphocytes, and potent agonists at CC chemokine receptor 1. *Blood* 1998; **91**:3118–26.
  - 54 Elagoz A, Henderson D, Babu PS *et al.* A truncated form of CKbeta8-1 is a potent agonist for human formyl peptide-receptor-like 1 receptor. *Br J Pharmacol* 2004; **141**:37–46.
  - 55 Nibbs RJ, Graham GJ. CCL27/PESKY: a novel paradigm for chemokine function. *Expert Opin Biol Ther* 2003; **3**:15–22.
  - 56 Baird JW, Nibbs RJ, Komai-Koma M *et al.* ESKine, a novel beta-chemokine, is differentially spliced to produce secretable and nuclear targeted isoforms. *J Biol Chem* 1999; **274**:33496–503.
  - 57 Ledee DR, Chen J, Tonelli LH, Takase H, Gery I, Zelenka PS. Differential expression of splice variants of chemokine CCL27 mRNA in lens, cornea, and retina of the normal mouse eye. *Mol Vis* 2004; **10**:663–7.
  - 58 Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; **3**:391–7.
  - 59 Hoh J, Ott J. Genetic dissection of diseases: design and methods. *Curr Opin Genet Dev* 2004; **14**:229–32.
  - 60 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6**:95–108.
  - 61 Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR, Kidd KK. Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. *Eur J Hum Genet* 2007; **15**:302–12.
  - 62 Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell* 2006; **126**:37–47.
  - 63 Stamm S, Ben-Ari S, Rafalska I *et al.* Function of alternative splicing. *Gene* 2005; **344**:1–20.
  - 64 Emi M, Keicho N, Tokunaga K *et al.* Association of diffuse panbronchiolitis with microsatellite polymorphism of the human interleukin 8 (IL-8) gene. *J Hum Genet* 1999; **44**:169–72.
  - 65 Vairaktaris E, Yapijakis C, Serefoglou Z *et al.* The interleukin-8 (-251A/T) polymorphism is associated with increased risk for oral squamous cell carcinoma. *Eur J Surg Oncol* 2006; **33**:504–7.
  - 66 Jiang ZD, DuPont HL, Garey K *et al.* A common polymorphism in the interleukin 8 gene promoter is associated with *Clostridium difficile* diarrhea. *Am J Gastroenterol* 2006; **101**:1112–6.
  - 67 Gyulai Z, Klausz G, Tiszai A *et al.* Genetic polymorphism of interleukin-8 (IL-8) is associated with *Helicobacter pylori*-induced duodenal ulcer. *Eur Cytokine Netw* 2004; **15**:353–8.
  - 68 Ohyauchi M, Imatani A, Yonechi M *et al.* The polymorphism interleukin 8–251 A/T influences the susceptibility of *Helicobacter pylori* related gastric diseases in the Japanese population. *Gut* 2005; **54**:330–5.
  - 69 Kato I, van Doorn LJ, Canzian F *et al.* Host–bacterial interaction in the development of gastric precancerous lesions in a high risk population for gastric cancer in Venezuela. *Int J Cancer* 2006; **119**:1666–71.
  - 70 Hofner P, Gyulai Z, Kiss ZF *et al.* Genetic polymorphisms of NOD1 and IL-8, but not polymorphisms of TLR4 genes, are associated with *Helicobacter pylori*-induced duodenal ulcer and gastritis. *Helicobacter* 2007; **12**:124–31.
  - 71 Hofner P, Balog A, Gyulai Z *et al.* Polymorphism in the IL-8 gene, but not in the TLR4 gene, increases the severity of acute pancreatitis. *Pancreatol* 2006; **6**:542–8.
  - 72 Lu W, Pan K, Zhang L, Lin D, Miao X, You W. Genetic polymorphisms of interleukin (IL)-1B, IL-1RN, IL-8, IL-10 and tumor necrosis factor {alpha} and risk of gastric cancer in a Chinese population. *Carcinogenesis* 2005; **26**:631–6.
  - 73 Shirai K, Ohmiya N, Taguchi A *et al.* Interleukin-8 gene polymorphism associated with susceptibility to non-cardiac gastric carcinoma with microsatellite instability. *J Gastroenterol Hepatol* 2006; **21**:1129–35.
  - 74 Landi S, Moreno V, Gioia-Patricola L *et al.* Association of common polymorphisms in inflammatory genes interleukin (IL) 6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer. *Cancer Res* 2003; **63**:3560–6.
  - 75 McCarron SL, Edwards S, Evans PR *et al.* Influence of cytokine gene polymorphisms on the development of prostate cancer. *Cancer Res* 2002; **62**:3369–72.
  - 76 Ross OA, O'Neill C, Rea IM *et al.* Functional promoter region polymorphism of the proinflammatory chemokine IL-8 gene associates with Parkinson's disease in the Irish. *Hum Immunol* 2004; **65**:340–6.
  - 77 Kamali-Sarvestani E, Nikseresht AR, Aliparasti MR, Vessal M. IL-8 (-251 A/T) and CXCR2 (+1208 C/T) gene polymorphisms and risk of multiple sclerosis in Iranian patients. *Neurosci Lett* 2006; **404**:159–62.
  - 78 Infante J, Llorca J, Berciano J, Combarros O. Interleukin-8, intercellular adhesion molecule-1 and tumour necrosis factor-alpha gene polymorphisms and the risk for multiple system atrophy. *J Neurol Sci* 2005; **228**:11–13.
  - 79 van der Kuyl AC, Polstra AM, Weverling GJ, Zorgdrager F, van den Burg R, Cornelissen M. An IL-8 gene promoter polymorphism is associated with the risk of the development of AIDS-related Kaposi's sarcoma: a case–control study. *Aids* 2004; **18**:1206–8.
  - 80 Lee EB, Kim JY, Zhao J, Park MH, Song YW. Haplotype association of IL-8 gene with Behcet's disease. *Tissue Antigens* 2007; **69**:128–32.

- 81 Galimberti D, Scalabrini D, Fenoglio C *et al.* CXCL10 haplotypes and multiple sclerosis: association and correlation with clinical course. *Eur J Neurol* 2007; **14**:162–7.
- 82 Coll B, Alonso-Villaverde C, Parra S *et al.* The stromal derived factor-1 mutated allele (SDF1-3'A) is associated with a lower incidence of atherosclerosis in HIV-infected patients. *Aids* 2005; **19**:1877–83.
- 83 Razmkhah M, Talei AR, Doroudchi M, Khalili-Azad T, Ghaderi A. Stromal cell-derived factor-1 (SDF-1) alleles and susceptibility to breast carcinoma. *Cancer Lett* 2005; **225**:261–6.
- 84 Razmkhah M, Doroudchi M, Ghayumi SM, Erfani N, Ghaderi A. Stromal cell-derived factor-1 (SDF-1) gene and susceptibility of Iranian patients with lung cancer. *Lung Cancer* 2005; **49**:311–5.
- 85 Zafropoulos A, Crikas N, Passam AM, Spandidos DA. Significant involvement of CCR2-64I and CXCL12-3a in the development of sporadic breast cancer. *J Med Genet* 2004; **41**:e59.
- 86 Dommange F, Cartron G, Espanel C *et al.* CXCL12 polymorphism and malignant cell dissemination/tissue infiltration in acute myeloid leukemia. *FASEB J* 2006; **20**:1913–5.
- 87 de Oliveira Cavassin GG, De Lucca FL, Delgado Andre N *et al.* Molecular investigation of the stromal cell-derived factor-1 chemokine in lymphoid leukemia and lymphoma patients from Brazil. *Blood Cells Mol Dis* 2004; **33**:90–3.
- 88 Gerli G, Vanelli C, Turri O *et al.* SDF1-3'A gene polymorphism is associated with chronic myeloproliferative disease and thrombotic events. *Clin Chem* 2005; **51**:2411–4.
- 89 Schroppel B, Fischereder M, Ashkar R *et al.* The impact of polymorphisms in chemokine and chemokine receptors on outcomes in liver transplantation. *Am J Transplant* 2002; **2**:640–5.
- 90 Ide A, Kawasaki E, Abiru N *et al.* Stromal-cell derived factor-1 chemokine gene variant is associated with type 1 diabetes age at onset in Japanese population. *Hum Immunol* 2003; **64**:973–8.
- 91 Takabatake N, Shibata Y, Abe S *et al.* A single nucleotide polymorphism in the CCL1 gene predicts acute exacerbations in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006; **174**:875–85.
- 92 Karrer S, Bosserhoff AK, Weiderer P *et al.* The -2518 promotor polymorphism in the MCP-1 gene is associated with systemic sclerosis. *J Invest Dermatol* 2005; **124**:92–8.
- 93 Szalai C, Kozma GT, Nagy A *et al.* Polymorphism in the gene regulatory region of MCP-1 is associated with asthma susceptibility and severity. *J Allergy Clin Immunol* 2001; **108**:375–81.
- 94 Keszei M, Nagy A, Kozma GT *et al.* Pediatric asthmatic patients have low serum levels of monocyte chemoattractant protein-1. *J Asthma* 2006; **43**:399–404.
- 95 Tucci M, Barnes EV, Sobel ES *et al.* Strong association of a functional polymorphism in the monocyte chemoattractant protein 1 promoter gene with lupus nephritis. *Arthritis Rheum* 2004; **50**:1842–9.
- 96 Ye DQ, Hu YS, Li XP *et al.* The correlation between monocyte chemoattractant protein-1 and the arthritis of systemic lupus erythematosus among Chinese. *Arch Dermatol Res* 2005; **296**:366–71.
- 97 Kim HL, Lee DS, Yang SH *et al.* The polymorphism of monocyte chemoattractant protein-1 is associated with the renal disease of SLE. *Am J Kidney Dis* 2002; **40**:1146–52.
- 98 Ozyurek AR, Gurses D, Ulger Z, Levent E, Bakiler AR, Berdeli A. Allelic frequency of the MCP-1 promoter -2518 polymorphism in the Turkish population and in Turkish patients with juvenile rheumatoid arthritis. *Clin Rheumatol* 2006; **26**:546–50.
- 99 Kruger B, Schroppel B, Ashkan R *et al.* A monocyte chemoattractant protein-1 (MCP-1) polymorphism and outcome after renal transplantation. *J Am Soc Nephrol* 2002; **13**:2585–9.
- 100 Ghilardi G, Biondi ML, La Torre A, Battaglioli L, Scorza R. Breast cancer progression and host polymorphisms in the chemokine system: role of the macrophage chemoattractant protein-1 (MCP-1) -2518 G allele. *Clin Chem* 2005; **51**:452–5.
- 101 Omori K, Kazama JJ, Song J *et al.* Association of the MCP-1 gene polymorphism A-2518G with carpal-tunnel syndrome in hemodialysis patients. *Amyloid* 2002; **9**:175–82.
- 102 Mori H, Kaneko Y, Narita I *et al.* Monocyte chemoattractant protein-1 A-2518G gene polymorphism and renal survival of Japanese patients with immunoglobulin A nephropathy. *Clin Exp Nephrol* 2005; **9**:297–303.
- 103 Wegscheider BJ, Weger M, Renner W *et al.* Role of the CCL2/MCP-1-2518A>G gene polymorphism in HLA-B27 associated uveitis. *Mol Vis* 2005; **11**:896–900.
- 104 Szalai C, Duba J, Prohászka Z *et al.* Involvement of polymorphisms in the chemokine system in the susceptibility for coronary artery disease (CAD). Coincidence of elevated Lp(a) and MCP-1-2518 G/G genotype in CAD patients. *Atherosclerosis* 2001; **158**:233–9.
- 105 Kim MP, Wahl LM, Yanek LR, Becker DM, Becker LC. A monocyte chemoattractant protein-1 gene polymorphism is associated with occult ischemia in a high-risk asymptomatic population. *Atherosclerosis* 2006; **193**:366–72.
- 106 McDermott DH, Yang Q, Kathiresan S *et al.* CCL2 polymorphisms are associated with serum monocyte chemoattractant protein-1 levels and myocardial infarction in the Framingham Heart Study. *Circulation* 2005; **112**:1113–20.
- 107 Pola R, Flex A, Gaetani E *et al.* Monocyte chemoattractant protein-1 (MCP-1) gene polymorphism and risk of Alzheimer's disease in Italians. *Exp Gerontol* 2004; **39**:1249–52.
- 108 Pae CU, Yu HS, Kim TS *et al.* Monocyte chemoattractant protein-1 (MCP1) promoter -2518 polymorphism may confer a susceptibility to major depressive disorder in the Korean population. *Psychiatry Res* 2004; **127**:279–81.
- 109 Yang B, Houlberg K, Millward A, Demaine A. Polymorphisms of chemokine and chemokine receptor genes in Type 1 diabetes mellitus and its complications. *Cytokine* 2004; **26**:114–21.
- 110 Simeoni E, Hoffmann MM, Winkelmann BR *et al.* Association between the A-2518G polymorphism in the monocyte chemoattractant protein-1 gene and insulin resistance and Type 2 diabetes mellitus. *Diabetologia* 2004; **47**:1574–80.
- 111 Flores-Villanueva PO, Ruiz-Morales JA, Song CH *et al.* A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. *J Exp Med* 2005; **202**:1649–58.
- 112 Ramasawmy R, Cunha-Neto E, Fae KC *et al.* The monocyte chemoattractant protein-1 gene polymorphism is associated with cardiomyopathy in human chagas disease. *Clin Infect Dis* 2006; **43**:305–11.
- 113 Park BL, Kim YJ, Cheong HS *et al.* Association of common promoter polymorphisms of MCP1 with hepatitis B virus clearance. *Exp Mol Med* 2006; **38**:694–702.
- 114 Vyshkina T, Shugart YY, Birnbaum G, Leist TP, Kalman B. Association of haplotypes in the beta-chemokine locus with multiple sclerosis. *Eur J Hum Genet* 2005; **13**:240–7.
- 115 Vyshkina T, Kalman B. Haplotypes within genes of beta-chemokines in 17q11 are associated with multiple sclerosis: a second phase study. *Hum Genet* 2005; **118**:67–75.

- 116 Modi WS, Goedert JJ, Strathdee S *et al.* MCP-1–MCP-3–Eotaxin gene cluster influences HIV-1 transmission. *Aids* 2003; **17**:2357–65.
- 117 Vyshkina T, Kalman B. Analyses of a MS-associated haplotype encompassing the CCL3 gene. *J Neuroimmunol* 2006; **176**:216–8.
- 118 Kim JJ, Lee JH, Jang CH *et al.* Chemokine RANTES promoter polymorphisms in allergic rhinitis. *Laryngoscope* 2004; **114**:666–9.
- 119 Leung TF, Tang NL, Lam CW *et al.* RANTES G-401A polymorphism is associated with allergen sensitization and FEV1 in Chinese children. *Respir Med* 2005; **99**:216–9.
- 120 Fryer AA, Spiteri MA, Bianco A *et al.* The -403 G→A promoter polymorphism in the RANTES gene is associated with atopy and asthma. *Genes Immun* 2000; **1**:509–14.
- 121 Al-Abdulhadi SA, Helms PJ, Main M, Smith O, Christie G. Preferential transmission and association of the -403 G→A promoter RANTES polymorphism with atopic asthma. *Genes Immun* 2005; **6**:24–30.
- 122 Ye DQ, Yang SG, Li XP *et al.* Polymorphisms in the promoter region of RANTES in Han Chinese and their relationship with systemic lupus erythematosus. *Arch Dermatol Res* 2005; **297**:108–13.
- 123 Makki RF, al Sharif F, Gonzalez-Gay MA, Garcia-Porrúa C, Ollier WE, Hajeer AH. RANTES gene polymorphism in polymyalgia rheumatica, giant cell arteritis and rheumatoid arthritis. *Clin Exp Rheumatol* 2000; **18**:391–3.
- 124 Wang CR, Guo HR, Liu MF. RANTES promoter polymorphism as a genetic risk factor for rheumatoid arthritis in the Chinese. *Clin Exp Rheumatol* 2005; **23**:379–84.
- 125 Gade-Andavolu R, Comings DE, MacMurray J *et al.* RANTES: a genetic risk marker for multiple sclerosis. *Mult Scler* 2004; **10**:536–9.
- 126 McDermott DH, Beecroft MJ, Kleeburger CA *et al.* Chemokine RANTES promoter polymorphism affects risk of both HIV infection and disease progression in the Multicenter AIDS Cohort Study. *Aids* 2000; **14**:2671–8.
- 127 Ahlenstiel G, Iwan A, Nattermann J *et al.* Distribution and effects of polymorphic RANTES gene alleles in HIV/HCV coinfection – a prospective cross-sectional study. *World J Gastroenterol* 2005; **11**:7631–8.
- 128 Hellier S, Frodsham AJ, Hennig BJ *et al.* Association of genetic variants of the chemokine receptor CCR5 and its ligands, RANTES and MCP-2, with outcome of HCV infection. *Hepatology* 2003; **38**:1468–76.
- 129 Promrat K, McDermott DH, Gonzalez CM *et al.* Associations of chemokine system polymorphisms with clinical outcomes and treatment responses of chronic hepatitis C. *Gastroenterology* 2003; **124**:352–60.
- 130 Takada T, Suzuki E, Ishida T *et al.* Polymorphism in RANTES chemokine promoter affects extent of sarcoidosis in a Japanese population. *Tissue Antigens* 2001; **58**:293–8.
- 131 Simeoni E, Winkelmann BR, Hoffmann MM *et al.* Association of RANTES G-403A gene polymorphism with increased risk of coronary arteriosclerosis. *Eur Heart J* 2004; **25**:1438–46.
- 132 Yamaguchi S, Yamada Y, Metoki N *et al.* Genetic risk for atherothrombotic cerebral infarction in individuals stratified by sex or conventional risk factors for atherosclerosis. *Int J Mol Med* 2006; **18**:871–83.
- 133 Boger CA, Fischereder M, Deinzer M *et al.* RANTES gene polymorphisms predict all-cause and cardiac mortality in type 2 diabetes mellitus hemodialysis patients. *Atherosclerosis* 2005; **183**:121–9.
- 134 Hizawa N, Yamaguchi E, Konno S, Tanino Y, Jinushi E, Nishimura M. A functional polymorphism in the RANTES gene promoter is associated with the development of late-onset asthma. *Am J Respir Crit Care Med* 2002; **166**:686–90.
- 135 Yao TC, Kuo ML, See LC *et al.* The RANTES promoter polymorphism: a genetic risk factor for near-fatal asthma in Chinese children. *J Allergy Clin Immunol* 2003; **111**:1285–92.
- 136 Nakajima K, Tanaka Y, Nomiya T *et al.* RANTES promoter genotype is associated with diabetic nephropathy in type 2 diabetic subjects. *Diabetes Care* 2003; **26**:892–8.
- 137 Liao CH, Yao TC, Chung HT, See LC, Kuo ML, Huang JL. Polymorphisms in the promoter region of RANTES and the regulatory region of monocyte chemoattractant protein-1 among Chinese children with systemic lupus erythematosus. *J Rheumatol* 2004; **31**:2062–7.
- 138 Cooke GS, Tosh K, Ramaley PA *et al.* A polymorphism that reduces RANTES expression is associated with protection from death in HIV-seropositive Ugandans with advanced disease. *J Infect Dis* 2006; **194**:666–9.
- 139 Zhernakova A, Alizadeh BZ, Eerligh P *et al.* Genetic variants of RANTES are associated with serum RANTES level and protection for type 1 diabetes. *Genes Immun* 2006; **7**:544–9.
- 140 Fiten P, Vandenbroeck K, Dubois B *et al.* Microsatellite polymorphisms in the gene promoter of monocyte chemotactic protein-3 and analysis of the association between monocyte chemotactic protein-3 alleles and multiple sclerosis development. *J Neuroimmunol* 1999; **95**:195–201.
- 141 Nelissen I, Dubois B, Goris A, Ronsse I, Carton H, Opendakker G, Gelatinase B, PECAM-1 and MCP-3 gene polymorphisms in Belgian multiple sclerosis. *J Neurol Sci* 2002; **200**:43–8.
- 142 Tsunemi Y, Saeki H, Nakamura K *et al.* Eotaxin gene single nucleotide polymorphisms in the promoter and exon regions are not associated with susceptibility to atopic dermatitis, but two of them in the promoter region are associated with serum IgE levels in patients with atopic dermatitis. *J Dermatol Sci* 2002; **29**:222–8.
- 143 Zee RY, Cook NR, Cheng S *et al.* Threonine for alanine substitution in the eotaxin (CCL11) gene and the risk of incident myocardial infarction. *Atherosclerosis* 2004; **175**:91–4.
- 144 Chae SC, Lee YC, Park YR *et al.* Analysis of the polymorphisms in eotaxin gene family and their association with asthma, IgE, and eosinophil. *Biochem Biophys Res Commun* 2004; **320**:131–7.
- 145 Park YR, Choi SC, Lee ST, Kim KS, Chae SC, Chung HT. The association of eotaxin-2 and eotaxin-3 gene polymorphisms in a Korean population with ulcerative colitis. *Exp Mol Med* 2005; **37**:553–8.
- 146 Chae SC, Park YR, Shim SC, Lee IK, Chung HT. Eotaxin-3 gene polymorphisms are associated with rheumatoid arthritis in a Korean population. *Hum Immunol* 2005; **66**:314–20.
- 147 Chae SC, Park YR, Oh GJ, Lee JH, Chung HT. The suggestive association of eotaxin-2 and eotaxin-3 gene polymorphisms in Korean population with allergic rhinitis. *Immunogenetics* 2005; **56**:760–4.
- 148 Shirozu M, Nakano T, Inazawa J *et al.* Structure and chromosomal localization of the human stromal cell-derived factor 1 (SDF1) gene. *Genomics* 1995; **28**:495–500.
- 149 Gortz A, Nibbs RJ, McLean P *et al.* The chemokine ESkin/CCL27 displays novel modes of intracrine and paracrine function. *J Immunol* 2002; **169**:1387–94.

## ***5. DISCUSSIÓ***



## 5. DISCUSSIÓ

Malgrat que en aquest apartat es presenta una discussió global dels resultats, per tal de facilitar la seva comprensió s'ha subdividit en subapartats corresponents als diferents capítols de resultats. La separació en subapartats tracta de clarificar els plantejaments discutits, intentant complementar-los sense repetir els conceptes ja desenvolupats en les discussions de cada article (sovint poc desenvolupats per les limitacions intrínseques a les publicacions).

### 5.1. Discussió del capítol I

CCL3 i CCL4 són dues quimiocines humanes que comparteixen múltiples característiques estructurals i funcionals. De fet, l'estreta relació entre ambdues es manté des del seu descobriment: el 1988 es va purificar un doblet de proteïnes amb activitat inflammatòria procedent del sobrenadant de macròfags murins estimulats amb endotoxina [83]. Degut a aquestes propietats inflammatòries, la barreja de proteïnes va ser anomenada MIP-1 (de l'anglès *macrophage inflammatory protein-1*). Estudis bioquímics posteriors van separar i caracteritzar els components d'aquest doblet, demostrant que estava format per dues proteïnes diferents, però altament relacionades, MIP-1 $\alpha$  i MIP1 $\beta$ , que compartien un 68% d'homologia aminoacídica [129]. Ràpidament es va trobar una elevada homologia de seqüència nucleotídica entre el cDNA del MIP-1 $\alpha$  murí i el cDNA humà corresponent al gen LD78 $\alpha$  o GOS19, clonat feia poc temps a partir de limfòcits estimulats [130, 131]. Conseqüentment es va assumir que LD78 $\alpha$  (l'actual CCL3) era l'homòleg humà del MIP-1 $\alpha$  murí. Pocs anys després (1990), i de forma inesperada, diferents grups van presentar evidències de l'existència d'un segon i, fins i tot, un tercer gen humà similar a LD78 $\alpha$  [103, 132, 133]. El segon gen similar a LD78 $\alpha$ , anomenat LD78 $\beta$  (l'actual CCL3L), presentava un 94% d'homologia de seqüència nucleotídica amb LD78 $\alpha$  i es transcrivia en limfòcits activats, mentre que el tercer gen similar a LD78 $\alpha$ , anomenat LD78 $\gamma$  (l'actual CCL3LP), semblava ser un pseudogen [105]. De forma similar al que succeí amb MIP-1 $\alpha$ , l'equivalent humà de MIP-1 $\beta$  va ser clonat independentment per diferents grups i va ser inicialment anomenat Act-2 o AT 744 (noms que corresponen a l'actual CCL4) [134-138]. De la mateixa manera, es va evidenciar l'existència d'un segon gen similar al MIP-1 $\beta$  humà, anomenat LAG-1 o AT 744.2 (noms que corresponen a l'actual CCL4L) [103]. En el cas de la parella LD78 $\alpha$  – LD78 $\beta$  (CCL3 – CCL3L), hi han hagut nombrosos estudis dedicats a caracteritzar aquestes dues variants i s'han detectat diferències funcionals interessants. Dels diferents treballs se'n desprèn que CCL3L presenta una activitat

quimioatracient superior a CCL3 respecte a monòcits i limfòcits; també s'ha demostrat que, en comparació a CCL3, CCL3L és un inhibidor més potent de la infecció pel virus HIV-1 R5 [139-141]. En el cas de la parella Act-2 – LAG-1 (CCL4 – CCL4L), després de la descripció inicial de l'existència d'ambdós gens, no es va aprofundir més en el coneixement del segon locus o *CCL4L*. Precisament, la primera part d'aquesta tesi doctoral es dedica bàsicament a l'estudi immunogenètic de les quimiocines CCL4 i CCL4L. El primer resultat interessant és l'evidència que els gens CCL4 i CCL4L expressen, a més a més dels trànscrips originalment descrits, variants originades per un fenomen de tall-i-unió alternatiu que exclou l'exó 2 de l'mRNA madur. Aquest fet no s'havia descrit fins al moment i planteja nous interrogants pel que fa a les funcions i regulació d'aquests nous trànscrips. Un dels aspectes que crida més l'atenció de la seqüència aminoacídica d'aquestes dues noves variants (anomenades CCL4 $\Delta$ 2 i CCL4L $\Delta$ 2) és la manca d'homologia amb CCL4 i CCL4L i, fins i tot, amb l'estructura bàsica de les quimiocines. Aquest fet es deu al canvi de pauta de lectura que es dona en unir-se els exons 1 i 3. La seqüència resultant de CCL4 $\Delta$ 2 i CCL4L $\Delta$ 2 conserva tan sols els 2 primers aminoàcids respecte CCL4 i CCL4L, i la resta de seqüència no hi manté cap relació. De fet, es perden 3 de les quatre cisteïnes que caracteritzen les quimiocines. Per tant, tècnicament, podríem considerar que CCL4 $\Delta$ 2 i CCL4L $\Delta$ 2 no són quimiocines, almenys des d'un punt de vista estructural. Si aquests trànscrips realment es tradueixen i si aquestes petites proteïnes (29 aminoàcids) mantindran algun tipus de capacitat quimioatracient és una incògnita, malgrat que *a priori* sembla difícil predir-ho. Per altra banda, la búsqueda de seqüències aminoacídiques de proteïnes conegudes amb una certa homologia a les proteïnes CCL4 $\Delta$ 2 i CCL4L $\Delta$ 2 no ha donat resultats rellevants.

L'estudi del locus *CCL4L* amb més profunditat també ha aportat la descoberta de noves variants que augmenten la complexitat d'aquesta quimiocina. En aquest cas l'efecte causant no és un fenomen de tall-i-unió alternatiu de l'mRNA, sinó un polimorfisme de nucleòtid únic (SNP, rs4796195) que es situa a la seqüència acceptora de tall del segon intró. Cal remarcar que els SNPs que afecten les seqüències que marquen el tall dels introns de l'mRNA primari són molt poc freqüents, ja que el seu efecte sobre la funcionalitat del gen acostuma a ser dramàtic [142-144]. Aquest cas no és una excepció: l'SNP converteix el dinucleòtid acceptor de tall universal AG present en l'intró 2 de la variant *CCL4L1*, en GG a la variant *CCL4L2*. Això fa que en la variant L2 desaparegui aquest lloc acceptor de tall original i, per tant, no es produeix cap dels dos trànscrips que es generen a partir de *CCL4L1*. En lloc seu, es produeixen una nova col·lecció de trànscrips fruit de l'ús de noves seqüències acceptores de tall que es troben al voltant de la original, ara mutada. Aquestes noves seqüències de tall s'anomenen críptiques, ja que originalment estan "camuflades" i s'utilitzen tan sols quan el lloc de tall natural queda desbaratat per efecte d'una mutació. S'ha observat de forma general que aquestes seqüències críptiques acostumen a situar-se prop de les autèntiques (com passa en el

nostre cas) [145]. Mitjançant anàlisis estadístics s'han creat algorismes que permeten quantificar la força que té una possible seqüència de tall per ser seleccionada i s'ha observat que les seqüències originals de tall normalment tenen una puntuació més elevada que les críptiques [146-149]. Això, però, no sempre és així, per la qual cosa es creu que hi ha altres elements del context que intervenen en el procés de selecció d'una seqüència de tall. Un exemple clar d'aquest fet és el que succeeix en el cas de *CCL4L2*: el lloc de tall seleccionat amb més freqüència (78,2%, origina la variant *CCL4L2*) no és precisament el que presenta una major puntuació teòrica. Malgrat tot, és el que està situat més aprop de la seqüència original i, de fet, origina una proteïna que mancada dels 5 primers aminoàcids de l'exó 3, però que manté el marc de lectura. La resta de seqüències acceptores de tall seleccionades (totes amb una freqüència inferior al 10%) originen trànscrips que, o bé donen lloc a una proteïna truncada (sense l'exó 3) o bé presenten un canvi de marc de lectura i donen lloc a proteïnes amb seqüències no relacionades amb *CCL4* o *CCL4L1* a partir de l'exó 3. Un aspecte interessant és que en la variant al·lèlica *CCL4L2*, igual com passa amb *CCL4* i *CCL4L1*, es manté la generació adicional de trànscrips originats per un fenomen de tall-i-unió alternatiu que exclou l'exó 2. Això es demostra amb la clonació de les formes *CCL4L2Δ2* i *CCL4L2bΔ2* (corresponents a *CCL4L2* i *CCL4L2b* sense el segon exó). De fet, possiblement existeixin tots els trànscrips generats a partir de *CCL4L2* en la seva forma alternativa sense el segon exó, però degut a la seva poca presència quantitativa, no els hem pogut clonar tots.

L'estudi de l'expressió, a nivell d'mRNA, de les variants de *CCL4* i *CCL4L* presenta algunes dades interessants. Quantificant alhora tots els trànscrips derivats de *CCL4* i *CCL4L* (utilitzant oligonucleòtids en zones comunes als exons 1 i 3) expressats per limfòcits T CD8 estimulats, s'observa per una banda que, independentment del genotip per a *CCL4L* (*L1L1*, *L1L2* i *L2L2*), es manté el pic d'expressió a les 6 hores post-estimulació, però s'evidencia que les mostres portadores de la variant al·lèlica *CCL4L2* produeixen menys quantitat global de missatger. Aquest fet es confirma quan analitzem quina part del total de trànscrips prové de cadascun dels dos gens: en mostres *L1L1*, aproximadament el 50% del mRNA prové del gen *CCL4L*, però en mostres *L2L2* aquesta proporció disminueix considerablement fins al 12%. Aquestes dades tenen una transcendència important ja que significa que en condicions normals (tenint en compte que la variant *L1* és àmpliament majoritària a la població caucàsica) la contribució dels locus *CCL4* i *CCL4L* en una situació inflammatòria és equivalent. Això situa les dues quimiocines a un nivell similar i reforça la importància de determinar les possibles diferències funcionals entre elles. És convenient recordar que els experiments d'expressió han estat realitzats a partir de limfòcits T CD8 estimulats amb PHA + IL-2 i que podria ser que en d'altres tipus cel·lulars, les proporcions exposades en aquest treball puguin variar. De fet, en un article publicat pocs mesos abans que el que conté aquestes dades, Lu *et al.* analitzen l'expressió de *CCL4* i *CCL4L* en monòcits i limfòcits B [150]. Les



conclusions generals d'aquest estudi mostren que les cèl·lules B (tant naïve com memòria) estimulades amb anti-IgM, expressen quantitats equivalents de CCL4 i CCL4L (com passa en el nostre cas amb els limfòcits T CD8). Però en el cas dels monòcits, expressen predominantment CCL4 (tot i que no mostren la relació numèrica). Malgrat l'interès dels resultats d'aquest estudi, cal remarcar que els autors no tenen en compte les dues variants al·lèliques principals de *CCL4L* originades per l'SNP rs4796195 (les quals ja hem vist que presenten diferències importants a nivell d'expressió). Per altra banda, desconeixen la presència de variants de *CCL4* i *CCL4L* originades per tall-i-unió alternatiu de l'mRNA. Curiosament, quan amplifiquen ambdós gens utilitzant primers situats als exons 1 i 3 els apareix, a part de la banda corresponent a CCL4 o CCL4L, una banda inesperada de menor pes molecular. Els autors comenten literalment "...these RT-PCR reactions of monocytes and stimulated B cells yielded a predominant band whose sequence was confirmed to be a CCL4 gene and a faster migrating band of lower, but variable, intensity. It is unclear what this lower band represents as it could not be sequenced". Sembla probable, doncs, que aquesta banda correspongui a les variants CCL4 $\Delta$ 2 i CCL4L $\Delta$ 2 (que, en el nostre estudi d'expressió, representen tan sols el 3% respecte els trànscrips de les variants completes CCL4 i CCL4L).

La repercussió del complex panorama genètic descrit per a *CCL4* i *CCL4L* requerirà posteriors estudis per a comprendre amb més detall com es regulen tots aquests processos. Una primera evidència, però, de la importància d'aquesta variabilitat la trobem en l'estudi cas-control que analitza la distribució de les variants originades per l'SNP rs4796195 (L1 i L2) en diferents patologies. Partint de la distribució de les freqüències al·lèliques i genotípiques en població sana (83,4% L1 vs 16,6% L2 i 70% L1L1, 27% L1L2 i 3% L2L2), s'ha determinat una distribució significativament diferent en el grup d'individus infectats pel virus HIV. Concretament, l'al·lel L2 està hiperrepresentat en el grup de pacients respecte els controls (28,6% vs 16,6%). Això indica que l'al·lel L2 representa un factor de susceptibilitat a contraure la infecció per HIV (sempre que hi hagi, òbviament, una exposició al virus). Una dada rellevant és la proporció d'individus homozigots L2L2 pràcticament es triplica en el grup d'infectats respecte els controls no infectats (3% vs 8,6%). Sembla clar que aquests resultats condueixen a la planificació d'estudis addicionals per veure el possible paper diferencial de les variants CCL4L1 i CCL4L2 respecte a la infecció pel virus HIV. De fet el nostre laboratori ja està treballant en aquesta direcció: la idea és produir de forma recombinant les variants principals de CCL4 i CCL4L per realitzar un ampli panell d'experiments funcionals (incloent estudis de bloqueig de la infecció per HIV) que ens permetin definir les diferències entre elles. La prioritat és començar per les proteïnes CCL4, CCL4L1 i CCL4L2, que són els productes quantitativament més importants dels gens *CCL4* i *CCL4L* (variants al·lèliques L1 i L2). Com ja s'ha comentat a la introducció, entre CCL4 i CCL4L1 tan sols existeix un aminoàcid de diferència (Ser<sub>47</sub> → Gly<sub>47</sub>) però que pot tenir conseqüències importants ja

que l'aminoàcid Gly<sub>47</sub> és incapaç de formar un pont d'hidrogen que estabilitza una nança de la zona implicada en la unió de CCL4 a GAGs i també en la formació de dímers d'aquesta quimiocina. En el cas de CCL4L2, el canvi és més important ja que li manquen els 5 primers aminoàcids de l'exó 3 (F<sub>42</sub>Q<sub>43</sub>T<sub>44</sub>K<sub>45</sub>R<sub>46</sub>) que, justament, precedeixen la glicina 47 que diferencia CCL4L1 respecte CCL4. En aquest cas, l'efecte sobre la zona implicada en la unió a GAGs és dràstic, ja que en un treball del 1999 Koopmann *et al* van demostrar de forma elegant que, dels 6 residus bàsics de la quimiocina, K<sub>45</sub> i R<sub>46</sub> són essencials per aquesta unió [84]. En aquest estudi es van realitzar 6 mutants de CCL4 (un per a cadascun dels aminoàcids bàsics de la quimiocina) on es substituïa l'aminoàcid bàsic per una alanina i s'analitzava la seva unió a l'heparina. En aquest treball també es demostra que els aminoàcids essencials per a la unió a GAGs, no estan implicats en la unió de CCL4 al receptor i en la subsegüent activació cel·lular. Concretament, el mutant d'R<sub>46</sub> es comporta igual que la proteïna normal en els assaig de quimiotaxi de limfòcits T. Això explica que les conclusions del treball publicat el 2004 per Howard *et al* [106], siguin que existeix una redundància funcional entre les quimiocines CCL4 i CCL4L. En aquest estudi es realitzen diversos experiments *in vitro* que inclouen el càlcul d'afinitat d'unió al receptor, quimiotaxis i bloqueig de la infecció per HIV. Sembla clar que per dilucidar les possibles diferències entre CCL4 i CCL4L1, cal dissenyar experiments que tinguin en compte les condicions *in vivo*, on la unió a GAGs és crítica per la funcionalitat d'aquestes quimiocines [151-153]. Aquesta consideració també serveix per CCL4L2 ja que, malgrat que l'estructura tridimensional de la zona d'unió a GAGs sembla que queda seriosament afectada, podria ser que mantingués certa capacitat d'unió al receptor i d'activació cel·lular (la zona d'unió al receptor es troba a l'extrem N-terminal, força allunyada de la zona d'unió a GAGs que varia entre CCL4, CCL4L1 i CCL4L2).

## 5.2. Discussió del capítol II

L'exhaustiva descripció de la complexitat genètica i transcripcional de CCL4 i CCL4L que s'ha realitzat en el treball del capítol I, juntament amb totes les possibles conseqüències funcionals que s'han comentat, ens fa adonar de la necessitat de ser rigorosos en els estudis que incloguin CCL4 i/o CCL4L. Com que molts d'aquests estudis contenen experiments basats en l'ús de proteïnes recombinants adquirides comercialment, cal confirmar que el producte que ens subministren és realment aquell que demanem. Aquest fet que, segons en quins casos, pot semblar una obvietat, en el cas de CCL4 i CCL4L pot arribar a ser un problema important. En el capítol II s'han presentat casos de confusions importants en les seqüències aminoacídiques de les quimiocines CCL4 i CCL4L d'algunes empreses que, si no són detectades, poden portar als investigadors a establir conclusions errònies en els seus estudis. Això constata que, per una banda, les empreses especialitzades han d'estar constantment actualitzant la informació dels seus productes i,

per l'altra, els investigadors han d'aprofundir en les característiques del producte que demanen, més enllà de la descripció superficial que moltes vegades apareix en els catàlegs o pàgines web.

### 5.3. Discussió del capítol III

A tot el panorama descrit anteriorment, cal afegir l'existència d'una CNV que afecta la regió cromosòmica que conté els gens *CCL4L* i *CCL3L*. La primera evidència de que aquests dos gens es troben en nombre variable de còpies al genoma humà apareix el 2002 [107], en el marc d'un interès creixent en la detecció i coneixement d'aquest tipus de variació estructural. Tal i com s'ha comentat a la introducció, les variacions en nombre de còpies són força freqüents al genoma humà [111, 112, 116, 120], i poden tenir conseqüències fenotípiques importants com a resultat de l'alteració de la dosi gènica, per interrupció de seqüències codificants o perturbant zones implicades en la regulació gènica [154, 155]. En els últims anys s'han desenvolupat múltiples mètodes per a la detecció de CNVs, que podem classificar de forma general en dos grans grups: a) els destinats a l'exploració del genoma de forma global (*genome-wide scans*) com són, per exemple, els mètodes basats en la hibridació genòmica comparada (*clone-based array-CGH*, *oligonucleotide-based array-CGH*) o en l'anàlisi d'SNPs (*SNP array*) i b) els destinats a analitzar regions concretes (*targeted scans*) com són, per exemple, els basats en la PCR a temps real o els recentment desenvolupats MLPA (*multiplex ligation-dependent probe amplification*) i MAPH (*multiplex amplification and probe hybridization*) [156, 157]. El nostre estudi (capítol III) utilitza la PCR a temps real per quantificar el nombre de còpies de *CCL4L* en les mostres del panell HGDP-CEPH, i el protocol concret utilitzat és una adaptació del que s'utilitza en treballs similars [108, 109, 158-160]. Els resultats demostren l'existència d'una elevada estructura poblacional pel que fa al nombre de còpies de *CCL4L* i el rang de variació del nombre de còpies del nostre panell de mostres va de 0 a 10 còpies de *CCL4L* per genoma diploide. Les poblacions de l'Àfrica Subsahariana són les que presenten un nombre més elevat de còpies de *CCL4L*, seguides de les d'Amèrica (central i del sud) i de les de l'est del continent asiàtic. A mesura que ens acostem a Europa, el nombre de còpies disminueix fins que, a l'arribar al nostre continent, els valors es situen, de mitjana, lleugerament per sota de 2 còpies/genoma diploide.

Una de les fites metodològiques d'aquest treball era com afrontar l'estudi dels SNPs en un gen amb variació en el nombre de còpies. Una de les preguntes que consideràvem importants poder respondre és: en una mostra determinada, quin nombre de còpies correspon a cadascuna de les variants al·lèliques de l'SNP? Òbviament aquesta qüestió no té cap dificultat de resposta en les mostres homozigotes per a qualsevol de les variants al·lèliques, però en el cas de mostres heterozigotes amb més de dues còpies, la

resposta no és evident. En el cas concret del gen *CCL4L*, aquesta situació pot arribar a complicar-se força per la presència de mostres heterozigotes amb 5, 6, 7 o més còpies (amb les possibles combinacions de distribució de còpies que implica). Com s'ha vist a l'article del capítol III, la solució ha estat realitzar la genotipificació utilitzant un sistema de PCR a temps real amb sondes FRET que, combinat amb el resultat de la quantificació de còpies realitzada anteriorment amb sondes TaqMan, ens ha permès, per a cada mostra, determinar el nombre de còpies de *CCL4L* corresponents a cadascun dels al·lels dels 2 SNPs (L1 o L2 en el cas de l'SNP rs4796195 / R o H en el cas de l'SNP).

La distribució poblacional dels 2 SNPs analitzats és molt diferent. L'SNP rs4796195 segueix un patró molt similar al descrit pel nombre de còpies: les poblacions de l'Àfrica Subsahariana i les d'Amèrica (central i del sud) són les que presenten una freqüència més elevada de l'al·lel menor L2 (41% i 43% respectivament), mentre que a mesura que ens acostem a Europa, la freqüència disminueix fins a situar-se en el 17% en aquest continent. Destacar que aquest 17% de mitjana de freqüència de l'al·lel L2 en població europea és molt similar al 16,6% obtingut en l'estudi poblacional del capítol I sobre 220 mostres procedent d'individus de població catalana. Pel que fa a l'SNP rs3744595, la distribució poblacional és totalment diferent. De fet, en aquest cas, l'al·lel menor H es troba pràcticament absent (freqüència de l'1%) en les poblacions de l'Àfrica Subsahariana i d'Amèrica (central i del sud) i presenta la màxima freqüència en poblacions de l'Orient Mitjà i Oceania (14% i 16% respectivament). Aquest comportament tan diferent dels 2 SNPs ens va sorprendre inicialment, ja que es troben força propers (tan sols separats per 647 nucleòtids) i, per tant, es podia esperar un cert desequilibri de lligament (LD) entre ells. Però un cop fet l'anàlisi s'observa la manca de LD que hi ha entre els 2 SNPs i que explica aquestes diferències. Tal i com s'ha comentat, la distribució poblacional de les freqüències de l'SNP rs4796195 recorda molt a la distribució del nombre de còpies de *CCL4L*. Aquest fet es confirma amb l'anàlisi de correlació entre les freqüències dels SNPs i el nombre de còpies de *CCL4L*. En el cas de rs4796195, la correlació amb la freqüència de l'al·lel menor i amb la freqüència d'heterozigosi és positiva i molt alta (0,731 i 0,890 respectivament). En canvi, l'SNP rs3744595 mostra una lleugera correlació negativa entre la freqüència de l'al·lel menor o la freqüència d'heterozigosi i el nombre de còpies de *CCL4L* (-0,491 i -0,186 respectivament). L'elevada correlació entre l'SNP rs4796195 i el nombre de còpies de *CCL4L* és un aspecte força interessant de l'estudi ja que ens permet usar un marcador tant fàcil de genotipar com un SNP, com a indicador del nombre aproximat de còpies d'una població concreta. En el cas de CNVs bial·lèliques, hi ha interès en trobar SNPs que hi estiguin en desequilibri de lligament, de tal manera que a través de genotipar l'SNP poguem determinar el genotip de la CNV en cada mostra. Això en el cas de CNVs multial·lèliques com la que afecta a *CCL4L* no és possible a nivell de mostres individuals,

però sí que en el cas de l'SNP rs4796195, ens pot servir per estimar de forma aproximada el nombre de còpies de *CCL4L* d'una població donada.

L'anàlisi de les freqüències fenotípiques dels SNPs també presenta alguns aspectes interessants a comentar, sobretot pel que fa a l'rs4796195. En el aquest cas d'aquest SNP trobem moltes poblacions que presenten freqüències d'heterozigosi superiors al 50%, que és el màxim esperat seguint l'equilibri Hardy-Weinberg (HWE), amb diverses poblacions que superen el 80% i arribant fins al cas extrem d'un 100% d'heterozigosi en una de les poblacions de l'Àfrica Subsahariana [*South East Africa (Bantu)*]. Aquest fet està causat per la variabilitat en nombre de còpies del gen *CCL4L*. Al realitzar aquest anàlisi hem considerat que les mostres heterozigotes són aquelles que presenten les dues variants al·lèliques, independentment del seu nombre de còpies. Per exemple, trobem diversos casos de mostres heterozigotes amb 6 còpies de *CCL4L* que es distribueixen al·lèlicament de forma tan diversa com: 5 còpies L1 / 1 còpia L2, 4 còpies L1 / 2 còpies L2 o 3 còpies L1 / 3 còpies L2. Això fa que les mostres heterozigotes formin un conjunt força heterogeni que, en segons quins casos, no hauria de ser tractat de forma unitària. Per exemple, si tenim en compte que l'al·lel L2 representa un factor de susceptibilitat a la infecció per HIV, en un estudi cas-control no hauríem de considerar les mostres heterozigotes com un sol grup (com es fa normalment en gens de còpia única), ja que, imaginant un cas extrem, podria tenir conseqüències molt diferents tenir una distribució de 5 còpies L1 / 1 còpia L2, o bé una de 1 còpia L1 / 5 còpies L2. En el cas del segon SNP (rs3744595), pràcticament no es produeixen desviacions del HWE ja que la seva freqüència és força més baixa (en especial a les regions que tenen un nombre més elevat de còpies de *CCL4L*). Aquesta característica desviació del HWE que es produeix en alguns SNPs que formen part de regions amb CNV, s'ha utilitzat justament com a eina per a la detecció d'aquestes CNVs.

Per tant, en estudis genètics de variació fenotípica que impliquin gens amb CNV, sembla important combinar la determinació del nombre de còpies amb la genotipificació d'SNPs rellevants, especialment en estudis de susceptibilitat/resistència a malaltia. En el cas de la relació de *CCL4L* en la infecció per HIV, s'han realitzat estudis que impliquen els tres elements analitzats en aquest treball (la CNV i els 2 SNPs) en la malaltia, de forma independent [109, 123, 128], però manquen estudis que combinin l'anàlisi conjunt. Aquest és un punt important que també tenim en compte de cara a estudis futurs. En el cas de les dades que hem obtingut en el panell HGDP-CEPH (cal recordar que són mostres de població normal), és curiós el fet que les poblacions que tenen un nombre més elevat de còpies de *CCL4L* (descriu com a factor de protecció vers la infecció per HIV [109]), també són les que presenten una freqüència més elevada de la variant al·lèlica *CCL4L2* (descriu com a factor de susceptibilitat vers la infecció per HIV [123]). A més a més, la manca de correlació de les variants R/H amb el nombre de còpies de *CCL4L*, augmenta encara més la complexitat d'interpretar el paper d'aquests 3 elements en la

infecció per HIV. Per tant, el balanç entre presentar un nombre més o menys elevat de còpies de *CCL4L-CCL3L* (recordar que parlem d'una regió cromosòmica que inclou ambdós gens) i la freqüència de l'al·lel L2, juntament amb tots els altres factors de susceptibilitat/resistència descrits per a la infecció per HIV, determinaran la contribució final de cada element en aquesta patologia.

#### 5.4. Discussió dels capítols IV i V

Com ja s'ha comentat, un dels objectius globals de la tesi ha sigut, més enllà de l'aprofundiment en el coneixement immunogenètic de *CCL4-CCL4L*, l'estudi global de la superfamília de les quimiocines. La gran quantitat de membres que la formen i la seva àmplia diversitat funcional, fan del seu estudi un viatge per la majoria d'elements que formen el sistema immunitari: des del coneixement dels diferents tipus cel·lulars que expressen o són dianes de l'acció de les quimiocines (que són pràcticament tots els tipus cel·lulars coneguts, transcendint més enllà dels que formen el sistema immunitari) fins als òrgans primaris o secundaris on les quimiocines són parcialment responsables de la l'organització interna dels diferents tipus cel·lulars que els formen (timus, medul·la òssia, ganglis...). Cal destacar que el caire immunogenètic del treball realitzat específicament amb *CCL4-CCL4L* ha influït en l'aproximació global a la superfamília de les quimiocines. De fet, la falta de publicacions que afrontessin la visió conjunta de la part genètica/genòmica de les quimiocines amb la seva funcionalitat general i amb el seu paper en malaltia, ens va fer plantejar la possibilitat de realitzar i publicar un estudi seriós des d'aquesta perspectiva. Finalment, degut a la gran quantitat d'informació recopilada i analitzada, el treball es va dividir en dues parts independents però relacionades que es van publicar en forma de revisió. Com que les pròpies publicacions són, en essència, discussions sobre els temes tractats, no es comenten en aquest apartat ja que considerem que seria redundant.



***6. CONCLUSIONS***





## 6. CONCLUSIONS

Les conclusions generals que podem extreure de la realització d'aquesta tesi doctoral són les següents:

1. Les quimiocines *CCL4* i *CCL4L* estan codificades per dos loci (*CCL4* i *CCL4L*) que generen dos tipus de trànscrips:
  - a.- els trànscrips complets, que corresponen a les quimiocines originalment descrites (*CCL4* i *CCL4L*).
  - b.- trànscrips originats per un fenomen de tall-i-unió alternatiu que elimina l'exó 2 de l'mRNA, generant les variants *CCL4 $\Delta$ 2* i *CCL4L $\Delta$ 2*.
2. El locus *CCL4L* presenta un SNP situat a la seqüència acceptora de tall de l'intró 2. La nova variant al·lèlica originada per aquest SNP (*CCL4L2*) presenta un patró d'expressió de trànscrips completament diferent al de la variant al·lèlica original (*CCL4L1*).
3. En cèl·lules T CD8 estimulades, el nivell d'expressió dels loci *CCL4* i *CCL4L* és equivalent quan es tracta de mostres homozigotes L1/L1. En el cas de mostres L2/L2, el percentatge de contribució del locus *CCL4L* respecte *CCL4* disminueix fins al 12%.
4. La freqüència de l'al·lel L2 es troba significativament augmentada en pacients HIV<sup>+</sup> respecte població sana.
5. Existeixen confusions en la seqüència aminoacídica de les quimiocines *CCL4* i *CCL4L* subministrades per algunes empreses.
6. La variabilitat en nombre de còpies del gen *CCL4L* presenta una peculiar estructura/distribució poblacional. Aquesta variabilitat ha estat estudiada en 1064 mostres d'individus corresponents a 51 poblacions mundials (panell HGDP-CEPH). Les poblacions de l'Àfrica Subsahariana són les que presenten un nombre de còpies de *CCL4L* més elevat (4,32 de mitjana) i les d'Europa les que en tenen menys (1,89 de mitjana).
7. És possible genotipar per PCR a temps real un determinat SNP en gens amb CNV, i definir exactament el nombre de còpies corresponents a cada variant al·lèlica.
8. En l'avaluació de SNPs de *CCL4L* relacionats amb la infecció per HIV, es demostren comportaments diferencials en la distribució poblacional mundial (s'han genotipat les mostres del panell HGDP-CEPH pels SNPs rs4796195 i rs3744595 de *CCL4L*). Per l'SNP rs4796195 les poblacions de l'Àfrica

Subsahariana i les d'Amèrica (central i del sud) són les que presenten una freqüència més elevada de l'al·lel menor L2 (41% i 43% respectivament) mentre que, a mesura que ens acostem a Europa, la freqüència disminueix fins a situar-se en el 17% en aquest continent. Per l'SNP rs3744595 la distribució poblacional és totalment diferent i en aquest cas, l'al·lel menor H es troba pràcticament absent en les poblacions de l'Àfrica Subsahariana i d'Amèrica (central i del sud) i presenta la màxima freqüència en poblacions de l'Orient Mitjà i Oceania (14% i 16% respectivament).

9. L'anàlisi de correlació entre les freqüències dels SNPs i el nombre de còpies de *CCL4L* mostra que, en el cas de l'SNP rs4796195, la correlació és positiva i molt alta (0,731 i 0,890 respectivament). En canvi, l'SNP rs3744595 mostra una lleugera correlació negativa entre la freqüència de l'al·lel menor o la freqüència d'heterozigosi i el nombre de còpies de *CCL4L* (-0,491 i -0,186 respectivament).
10. L'organització genòmica dels membres de la superfamília de les quimiocines està relacionada amb les seves funcions, ja que els grups de gens de quimiocines (o *clusters*) comparteixen funcions bàsiques.
11. Els polimorfismes i fenòmens de tall-i-unió alternatiu de l'mRNA augmenten el nombre de variants de les quimiocines i poden tenir influència en la susceptibilitat/resistència a diverses malalties.

## ***7. REFERÈNCIES***



## 7. REFERÈNCIES

1. Tedder TF, Steeber DA, Chen A and Engel P. The selectins: vascular adhesion molecules. *Faseb J* 1995;9:866-73
2. Gangenahalli GU, Singh VK, Verma YK, et al. Hematopoietic stem cell antigen CD34: role in adhesion or homing. *Stem Cells Dev* 2006;15:305-13
3. Hwang ST, Singer MS, Giblin PA, et al. GlyCAM-1, a physiologic ligand for L-selectin, activates beta 2 integrins on naive peripheral lymphocytes. *J Exp Med* 1996;184:1343-8
4. McEver RP, Cummings RD. Perspectives series: cell adhesion in vascular biology. Role of PSGL-1 binding to selectins in leukocyte recruitment. *J Clin Invest* 1997;100:485-91
5. Gille J, Swerlick RA. Integrins: role in cell adhesion and communication. *Ann N Y Acad Sci* 1996;797:93-106
6. Gahmberg CG. Leukocyte adhesion: CD11/CD18 integrins and intercellular adhesion molecules. *Curr Opin Cell Biol* 1997;9:643-50
7. Watt FM. Role of integrins in regulating epidermal adhesion, growth and differentiation. *Embo J* 2002;21:3919-26
8. Simmons DL. The role of ICAM expression in immunity and disease. *Cancer Surv* 1995;24:141-55
9. Hayflick JS, Kilgannon P and Gallatin WM. The intercellular adhesion molecule (ICAM) family of proteins. New members and novel functions. *Immunol Res* 1998;17:313-27
10. Cook-Mills JM. VCAM-1 signals during lymphocyte migration: role of reactive oxygen species. *Mol Immunol* 2002;39:499-508
11. Fong S, Jones S, Renz ME, et al. Mucosal addressin cell adhesion molecule-1 (MAdCAM-1). Its binding motif for alpha 4 beta 7 and role in experimental colitis. *Immunol Res* 1997;16:299-311

12. Miyasaka M, Tanaka T. Lymphocyte trafficking across high endothelial venules: dogmas and enigmas. *Nat Rev Immunol* 2004;4:360-70
13. Zlotnik A, Yoshie O. Chemokines: a new classification system and their role in immunity. *Immunity* 2000;12:121-7
14. Kelner GS, Kennedy J, Bacon KB, et al. Lymphotactin: a cytokine that represents a new class of chemokine. *Science* 1994;266:1395-9
15. Bazan JF, Bacon KB, Hardiman G, et al. A new class of membrane-bound chemokine with a CX3C motif. *Nature* 1997;385:640-4
16. Miller MD, Wilson SD, Dorf ME, Seuanez HN, O'Brien SJ and Krangel MS. Sequence and chromosomal location of the I-309 gene. Relationship to genes encoding a family of inflammatory cytokines. *J Immunol* 1990;145:2737-44
17. Schulz-Knappe P, Magert HJ, Dewald B, et al. HCC-1, a novel chemokine from human plasma. *J Exp Med* 1996;183:295-9
18. Nagira M, Imai T, Hieshima K, et al. Molecular cloning of a novel human CC chemokine secondary lymphoid-tissue chemokine that is a potent chemoattractant for lymphocytes and mapped to chromosome 9p13. *J Biol Chem* 1997;272:19518-24
19. Youn BS, Zhang SM, Lee EK, et al. Molecular cloning of leukotactin-1: a novel human beta-chemokine, a chemoattractant for neutrophils, monocytes, and lymphocytes, and a potent agonist at CC chemokine receptors 1 and 3. *J Immunol* 1997;159:5201-5
20. White JR, Imburgia C, Dul E, et al. Cloning and functional characterization of a novel human CC chemokine that binds to the CCR3 receptor and activates human eosinophils. *J Leukoc Biol* 1997;62:667-75
21. Skelton NJ, Aspiras F, Ogez J and Schall TJ. Proton NMR assignments and solution conformation of RANTES, a chemokine of the C-C type. *Biochemistry* 1995;34:5329-42

22. Gong JH, Clark-Lewis I. Antagonists of monocyte chemoattractant protein 1 identified by modification of functionally critical NH<sub>2</sub>-terminal residues. *J Exp Med* 1995;181:631-40
23. Witt DP, Lander AD. Differential binding of chemokines to glycosaminoglycan subpopulations. *Curr Biol* 1994;4:394-400
24. Covell DG, Smythers GW, Gronenborn AM and Clore GM. Analysis of hydrophobicity in the alpha and beta chemokine families and its relevance to dimerization. *Protein Sci* 1994;3:2064-72
25. Horcher M, Rot A, Aschauer H and Besemer J. IL-8 derivatives with a reduced potential to form homodimers are fully active in vitro and in vivo. *Cytokine* 1998;10:1-12
26. Laurence JS, LiWang AC and LiWang PJ. Effect of N-terminal truncation and solution conditions on chemokine dimer stability: nuclear magnetic resonance structural analysis of macrophage inflammatory protein 1 beta mutants. *Biochemistry* 1998;37:9346-54
27. Morimoto C, Schlossman SF. The structure and function of CD26 in the T-cell immune response. *Immunol Rev* 1998;161:55-70
28. De Meester I, Korom S, Van Damme J and Scharpe S. CD26, let it cut or cut it down. *Immunol Today* 1999;20:367-75
29. Van Damme J, Struyf S, Wuyts A, et al. The role of CD26/DPP IV in chemokine processing. *Chem Immunol* 1999;72:42-56
30. Proost P, De Meester I, Schols D, et al. Amino-terminal truncation of chemokines by CD26/dipeptidyl-peptidase IV. Conversion of RANTES into a potent inhibitor of monocyte chemotaxis and HIV-1-infection. *J Biol Chem* 1998;273:7222-7
31. Proost P, Struyf S, Schols D, et al. Truncation of macrophage-derived chemokine by CD26/ dipeptidyl-peptidase IV beyond its predicted cleavage site affects chemotactic activity and CC chemokine receptor 4 interaction. *J Biol Chem* 1999;274:3988-93



32. Lambeir AM, Proost P, Durinx C, et al. Kinetic investigation of chemokine truncation by CD26/dipeptidyl peptidase IV reveals a striking selectivity within the chemokine family. *J Biol Chem* 2001;276:29839-45
33. Struyf S, Menten P, Lenaerts JP, et al. Diverging binding capacities of natural LD78beta isoforms of macrophage inflammatory protein-1alpha to the CC chemokine receptors 1, 3 and 5 affect their anti-HIV-1 activity and chemotactic potencies for neutrophils and eosinophils. *Eur J Immunol* 2001;31:2170-8
34. Proost P, Menten P, Struyf S, Schutyser E, De Meester I and Van Damme J. Cleavage by CD26/dipeptidyl peptidase IV converts the chemokine LD78beta into a most efficient monocyte attractant and CCR1 agonist. *Blood* 2000;96:1674-80
35. Clark-Lewis I, Dewald B, Loetscher M, Moser B and Baggiolini M. Structural requirements for interleukin-8 function identified by design of analogs and CXC chemokine hybrids. *J Biol Chem* 1994;269:16075-81
36. Schraufstatter IU, Barritt DS, Ma M, Oades ZG and Cochrane CG. Multiple sites on IL-8 responsible for binding to alpha and beta IL-8 receptors. *J Immunol* 1993;151:6418-28
37. Keane MP, Arenberg DA, Moore BB, Addison CL and Strieter RM. CXC chemokines and angiogenesis/angiostasis. *Proc Assoc Am Physicians* 1998;110:288-96
38. Gengrinovitch S, Greenberg SM, Cohen T, et al. Platelet factor-4 inhibits the mitogenic activity of VEGF121 and VEGF165 using several concurrent mechanisms. *J Biol Chem* 1995;270:15059-65
39. Tanaka T, Manome Y, Wen P, Kufe DW and Fine HA. Viral vector-mediated transduction of a modified platelet factor 4 cDNA inhibits angiogenesis and tumor growth. *Nat Med* 1997;3:437-42
40. Cole KE, Strick CA, Paradis TJ, et al. Interferon-inducible T cell alpha chemoattractant (I-TAC): a novel non-ELR CXC chemokine with potent activity on activated T cells through selective high affinity binding to CXCR3. *J Exp Med* 1998;187:2009-21

41. Tachibana K, Hirota S, Iizasa H, et al. The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature* 1998;393:591-4
42. Bleul CC, Fuhlbrigge RC, Casasnovas JM, Aiuti A and Springer TA. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J Exp Med* 1996;184:1101-9
43. Nagasawa T, Nakajima T, Tachibana K, et al. Molecular cloning and characterization of a murine pre-B-cell growth-stimulating factor/stromal cell-derived factor 1 receptor, a murine homolog of the human immunodeficiency virus 1 entry coreceptor fusin. *Proc Natl Acad Sci U S A* 1996;93:14726-9
44. Aiuti A, Webb IJ, Bleul C, Springer T and Gutierrez-Ramos JC. The chemokine SDF-1 is a chemoattractant for human CD34+ hematopoietic progenitor cells and provides a new mechanism to explain the mobilization of CD34+ progenitors to peripheral blood. *J Exp Med* 1997;185:111-20
45. Zou YR, Kottmann AH, Kuroda M, Taniuchi I and Littman DR. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature* 1998;393:595-9
46. Amara A, Gall SL, Schwartz O, et al. HIV coreceptor downregulation as antiviral principle: SDF-1alpha-dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication. *J Exp Med* 1997;186:139-46
47. Dealwis C, Fernandez EJ, Thompson DA, Simon RJ, Siani MA and Lolis E. Crystal structure of chemically synthesized [N33A] stromal cell-derived factor 1alpha, a potent ligand for the HIV-1 "fusin" coreceptor. *Proc Natl Acad Sci U S A* 1998;95:6941-6
48. Gunn MD, Ngo VN, Ansel KM, Eklund EH, Cyster JG and Williams LT. A B-cell-homing chemokine made in lymphoid follicles activates Burkitt's lymphoma receptor-1. *Nature* 1998;391:799-803

49. Gunn MD, Tangemann K, Tam C, Cyster JG, Rosen SD and Williams LT. A chemokine expressed in lymphoid high endothelial venules promotes the adhesion and chemotaxis of naive T lymphocytes. *Proc Natl Acad Sci U S A* 1998;95:258-63
50. Legler DF, Loetscher M, Roos RS, Clark-Lewis I, Baggiolini M and Moser B. B cell-attracting chemokine 1, a human CXC chemokine expressed in lymphoid tissues, selectively attracts B lymphocytes via BLR1/CXCR5. *J Exp Med* 1998;187:655-60
51. Berkhout TA, Sarau HM, Moores K, et al. Cloning, in vitro expression, and functional characterization of a novel human CC chemokine of the monocyte chemoattractant protein (MCP) family (MCP-4) that binds and signals through the CC chemokine receptor 2B. *J Biol Chem* 1997;272:16404-13
52. Luster AD, Rothenberg ME. Role of the monocyte chemoattractant protein and eotaxin subfamily of chemokines in allergic inflammation. *J Leukoc Biol* 1997;62:620-33
53. Sauder DN, Mounessa NL, Katz SI, Dinarello CA and Gallin JI. Chemotactic cytokines: the role of leukocytic pyrogen and epidermal cell thymocyte-activating factor in neutrophil chemotaxis. *J Immunol* 1984;132:828-32
54. Yoshimura T, Matsushima K, Tanaka S, et al. Purification of a human monocyte-derived neutrophil chemotactic factor that has peptide sequence similarity to other host defense cytokines. *Proc Natl Acad Sci U S A* 1987;84:9233-7
55. Salcedo R, Young HA, Ponce ML, et al. Eotaxin (CCL11) induces in vivo angiogenic responses by human CCR3+ endothelial cells. *J Immunol* 2001;166:7571-8
56. Roos RS, Loetscher M, Legler DF, Clark-Lewis I, Baggiolini M and Moser B. Identification of CCR8, the receptor for the human CC chemokine I-309. *J Biol Chem* 1997;272:17251-4
57. Zingoni A, Soto H, Hedrick JA, et al. The chemokine receptor CCR8 is preferentially expressed in Th2 but not Th1 cells. *J Immunol* 1998;161:547-51

58. Tiffany HL, Lautens LL, Gao JL, et al. Identification of CCR8: a human monocyte and thymus receptor for the CC chemokine I-309. *J Exp Med* 1997;186:165-70
59. Schutysen E, Richmond A and Van Damme J. Involvement of CC chemokine ligand 18 (CCL18) in normal and pathological processes. *J Leukoc Biol* 2005;78:14-26
60. Tedla N, Wang HW, McNeil HP, et al. Regulation of T lymphocyte trafficking into lymph nodes during an immune response by the chemokines macrophage inflammatory protein (MIP)-1 alpha and MIP-1 beta. *J Immunol* 1998;161:5663-72
61. Dairaghi DJ, Franz-Bacon K, Callas E, et al. Macrophage inflammatory protein-1beta induces migration and activation of human thymocytes. *Blood* 1998;91:2905-13
62. Taub DD, Turcovski-Corrales SM, Key ML, Longo DL and Murphy WJ. Chemokines and T lymphocyte activation: I. Beta chemokines costimulate human T lymphocyte activation in vitro. *J Immunol* 1996;156:2095-2103
63. Adams DH, Lloyd AR. Chemokines: leucocyte recruitment and activation cytokines. *Lancet* 1997;349:490-5
64. Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC and Lusso P. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* 1995;270:1811-5
65. Cocchi F, DeVico AL, Yarchoan R, et al. Higher macrophage inflammatory protein (MIP)-1alpha and MIP-1beta levels from CD8+ T cells are associated with asymptomatic HIV-1 infection. *Proc Natl Acad Sci U S A* 2000;97:13812-7
66. Pardigol A, Forssmann U, Zucht HD, et al. HCC-2, a human chemokine: gene structure, expression pattern, and biological activity. *Proc Natl Acad Sci U S A* 1998;95:6308-13
67. Hedrick JA, Helms A, Vicari A and Zlotnik A. Characterization of a novel CC chemokine, HCC-4, whose expression is increased by interleukin-10. *Blood* 1998;91:4242-7

68. Murphy PM. The molecular biology of leukocyte chemoattractant receptors. *Annu Rev Immunol* 1994;12:593-633
69. Premack BA, Schall TJ. Chemokine receptors: gateways to inflammation and infection. *Nat Med* 1996;2:1174-8
70. Yoshie O, Imai T and Nomiyama H. Novel lymphocyte-specific CC chemokines and their receptors. *J Leukoc Biol* 1997;62:634-44
71. Locati M, Murphy PM. Chemokines and chemokine receptors: biology and clinical relevance in inflammation and AIDS. *Annu Rev Med* 1999;50:425-40
72. Zlotnik A, Morales J and Hedrick JA. Recent advances in chemokines and chemokine receptors. *Crit Rev Immunol* 1999;19:1-47
73. Kehrl JH. Chemoattractant receptor signaling and the control of lymphocyte migration. *Immunol Res* 2006;34:211-27
74. Busillo JM, Benovic JL. Regulation of CXCR4 signaling. *Biochim Biophys Acta* 2007;1768:952-63
75. McFadden G, Lalani A, Everett H, Nash P and Xu X. Virus-encoded receptors for cytokines and chemokines. *Semin Cell Dev Biol* 1998;9:359-68
76. Pease JE, Murphy PM. Microbial corruption of the chemokine system: an expanding paradigm. *Semin Immunol* 1998;10:169-78
77. Bodaghi B, Jones TR, Zipeto D, et al. Chemokine sequestration by viral chemoreceptors as a novel viral escape strategy: withdrawal of chemokines from the environment of cytomegalovirus-infected cells. *J Exp Med* 1998;188:855-66
78. Vieira J, Schall TJ, Corey L and Geballe AP. Functional analysis of the human cytomegalovirus US28 gene by insertion mutagenesis with the green fluorescent protein gene. *J Virol* 1998;72:8158-65

79. Arvanitakis L, Geras-Raaka E, Varma A, Gershengorn MC and Cesarman E. Human herpesvirus KSHV encodes a constitutively active G-protein-coupled receptor linked to cell proliferation. *Nature* 1997;385:347-50
80. Lodi PJ, Garrett DS, Kuszewski J, et al. High-resolution solution structure of the beta chemokine hMIP-1 beta by multidimensional NMR. *Science* 1994;263:1762-7
81. Clements JM, Craig S, Gearing AJ, et al. Biological and structural properties of MIP-1 alpha expressed in yeast. *Cytokine* 1992;4:76-82
82. Patel SR, Evans S, Dunne K, et al. Characterization of the quaternary structure and conformational properties of the human stem cell inhibitor protein LD78 in solution. *Biochemistry* 1993;32:5466-71
83. Wolpe SD, Davatelis G, Sherry B, et al. Macrophages secrete a novel heparin-binding protein with inflammatory and neutrophil chemokinetic properties. *J Exp Med* 1988;167:570-81
84. Koopmann W, Ediriwickrema C and Krangel MS. Structure and function of the glycosaminoglycan binding site of chemokine macrophage-inflammatory protein-1 beta. *J Immunol* 1999;163:2120-7
85. Ward SG, Bacon K and Westwick J. Chemokines and T lymphocytes: more than an attraction. *Immunity* 1998;9:1-11
86. Ward SG, Westwick J. Chemokines: understanding their role in T-lymphocyte biology. *Biochem J* 1998;333 ( Pt 3):457-70
87. Bernardini G, Hedrick J, Sozzani S, et al. Identification of the CC chemokines TARC and macrophage inflammatory protein-1 beta as novel functional ligands for the CCR8 receptor. *Eur J Immunol* 1998;28:582-8
88. Garlisi CG, Xiao H, Tian F, et al. The assignment of chemokine-chemokine receptor pairs: TARC and MIP-1 beta are not ligands for human CC-chemokine receptor 8. *Eur J Immunol* 1999;29:3210-5

89. Mueller A, Mahmoud NG and Strange PG. Diverse signalling by different chemokines through the chemokine receptor CCR5. *Biochem Pharmacol* 2006;72:739-48
90. Mueller A, Strange PG. The chemokine receptor, CCR5. *Int J Biochem Cell Biol* 2004;36:35-8
91. Samson M, Labbe O, Mollereau C, Vassart G and Parmentier M. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry* 1996;35:3362-7
92. Gong W, Howard OM, Turpin JA, et al. Monocyte chemotactic protein-2 activates CCR5 and blocks CD4/CCR5-mediated HIV-1 entry/replication. *J Biol Chem* 1998;273:4289-92
93. Ruffing N, Sullivan N, Sharmeen L, Sodroski J and Wu L. CCR5 has an expanded ligand-binding repertoire and is the primary receptor used by MCP-2 on activated T cells. *Cell Immunol* 1998;189:160-8
94. Blanpain C, Migeotte I, Lee B, et al. CCR5 binds multiple CC-chemokines: MCP-3 acts as a natural antagonist. *Blood* 1999;94:1899-905
95. Hariharan D, Douglas SD, Lee B, Lai JP, Campbell DE and Ho WZ. Interferon-gamma upregulates CCR5 expression in cord and adult blood mononuclear phagocytes. *Blood* 1999;93:1137-44
96. Patterson BK, Czerniewski M, Andersson J, et al. Regulation of CCR5 and CXCR4 expression by type 1 and type 2 cytokines: CCR5 expression is downregulated by IL-10 in CD4-positive lymphocytes. *Clin Immunol* 1999;91:254-62
97. Huang Y, Paxton WA, Wolinsky SM, et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med* 1996;2:1240-3
98. Rollins BJ. Chemokines. *Blood* 1997;90:909-28
99. Bystry RS, Aluvihare V, Welch KA, Kallikourdis M and Betz AG. B cells and professional APCs recruit regulatory T cells via CCL4. *Nat Immunol* 2001;2:1126-32

100. Joosten SA, van Meijgaarden KE, Savage ND, et al. Identification of a human CD8+ regulatory T cell subset that mediates suppression through the chemokine CC chemokine ligand 4. *Proc Natl Acad Sci U S A* 2007;104:8029-34
101. Broxmeyer HE, Sherry B, Cooper S, et al. Macrophage inflammatory protein (MIP)-1 beta abrogates the capacity of MIP-1 alpha to suppress myeloid progenitor cell growth. *J Immunol* 1991;147:2586-94
102. Siveke JT, Hamann A. T helper 1 and T helper 2 cells respond differentially to chemokines. *J Immunol* 1998;160:550-4
103. Irving SG, Zipfel PF, Balke J, et al. Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q. *Nucleic Acids Res* 1990;18:3261-70
104. Wilson SD, Billings PR, D'Eustachio P, et al. Clustering of cytokine genes on mouse chromosome 11. *J Exp Med* 1990;171:1301-14
105. Hirashima M, Ono T, Nakao M, et al. Nucleotide sequence of the third cytokine LD78 gene and mapping of all three LD78 gene loci to human chromosome 17. *DNA Seq* 1992;3:203-12
106. Howard OM, Turpin JA, Goldman R and Modi WS. Functional redundancy of the human CCL4 and CCL4L1 chemokine genes. *Biochem Biophys Res Commun* 2004;320:927-31
107. Townson JR, Barcellos LF and Nibbs RJ. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol* 2002;32:3016-26
108. McKinney C, Merriman ME, Chapman PT, et al. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis* 2007
109. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 2005;307:1434-40



110. Scherer SW, Lee C, Birney E, et al. Challenges and standards in integrating surveys of structural variation. *Nat Genet* 2007;39:S7-15
111. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949-51
112. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science* 2004;305:525-8
113. de Vries BB, Pfundt R, Leisink M, et al. Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 2005;77:606-16
114. Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM and Nordenskjold M. Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). *J Med Genet* 2005;42:699-705
115. Sharp AJ, Locke DP, McGrath SD, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005;77:78-88
116. Tuzun E, Sharp AJ, Bailey JA, et al. Fine-scale structural variation of the human genome. *Nat Genet* 2005;37:727-32
117. Tyson C, Harvard C, Locker R, et al. Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH. *Am J Med Genet A* 2005;139:173-85
118. Conrad DF, Andrews TD, Carter NP, Hurles ME and Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 2006;38:75-81
119. Hinds DA, Stuve LL, Nilsen GB, et al. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072-9
120. McCarroll SA, Hadnott TN, Perry GH, et al. Common deletion polymorphisms in the human genome. *Nat Genet* 2006;38:86-92

121. Shao W, Tang J, Song W, et al. CCL3L1 and CCL4L1: variable gene copy number in adolescents with and without human immunodeficiency virus type 1 (HIV-1) infection. *Genes Immun* 2007;8:224-31
122. Ashhab Y, Dominguez O, Sospedra M, Roura-Mir C, Lucas-Martin A and Pujol-Borrell R. A one-tube polymerase chain reaction protocol demonstrates CC chemokine overexpression in Graves' disease glands. *J Clin Endocrinol Metab* 1999;84:2873-82
123. Colobran R, Adreani P, Ashhab Y, et al. Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. *J Immunol* 2005;174:5655-64
124. Cann HM, de Toma C, Cazes L, et al. A human genome diversity cell line panel. *Science* 2002;296:261-2
125. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science* 2002;298:2381-5
126. Conrad DF, Jakobsson M, Coop G, et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 2006;38:1251-60
127. Santos-Lopes SS, Pereira RW, Wilson IJ and Pena SD. A worldwide phylogeography for the human X chromosome. *PLoS ONE* 2007;2:e557
128. Capoulade-Metay C, Meyer L, Tran T, et al. Influence of the R22H variant of macrophage inflammatory protein 1beta/Lag-1 in HIV-1 survival. *Aids* 2005;19:831-3
129. Sherry B, Tekamp-Olson P, Gallegos C, et al. Resolution of the two components of macrophage inflammatory protein 1, and cloning and characterization of one of those components, macrophage inflammatory protein 1 beta. *J Exp Med* 1988;168:2251-9
130. Forsdyke DR. cDNA cloning of mRNAs which increase rapidly in human lymphocytes cultured with concanavalin-A and cycloheximide. *Biochem Biophys Res Commun* 1985;129:619-25

131. Obaru K, Fukuda M, Maeda S and Shimada K. A cDNA clone used to study mRNA inducible in human tonsillar lymphocytes by a tumor promoter. *J Biochem (Tokyo)* 1986;99:885-94
132. Blum S, Forsdyke RE and Forsdyke DR. Three human homologs of a murine gene encoding an inhibitor of stem cell proliferation. *DNA Cell Biol* 1990;9:589-602
133. Nakao M, Nomiya H and Shimada K. Structures of human genes coding for cytokine LD78 and their expression. *Mol Cell Biol* 1990;10:3646-58
134. Lipes MA, Napolitano M, Jeang KT, Chang NT and Leonard WJ. Identification, cloning, and characterization of an immune activation gene. *Proc Natl Acad Sci U S A* 1988;85:9704-8
135. Zipfel PF, Balke J, Irving SG, Kelly K and Siebenlist U. Mitogenic activation of human T cells induces two closely related genes which share structural similarities with a new family of secreted factors. *J Immunol* 1989;142:1582-90
136. Chang HC, Reinherz EL. Isolation and characterization of a cDNA encoding a putative cytokine which is induced by stimulation via the CD2 structure on human T lymphocytes. *Eur J Immunol* 1989;19:1045-51
137. Miller MD, Hata S, De Waal Malefyt R and Krangel MS. A novel polypeptide secreted by activated human T lymphocytes. *J Immunol* 1989;143:2907-16
138. Baixeras E, Roman-Roman S, Jitsukawa S, et al. Cloning and expression of a lymphocyte activation gene (LAG-1). *Mol Immunol* 1990;27:1091-102
139. Menten P, Struyf S, Schutyser E, et al. The LD78beta isoform of MIP-1alpha is the most potent CCR5 agonist and HIV-1-inhibiting chemokine. *J Clin Invest* 1999;104:R1-5
140. Nibbs RJ, Yang J, Landau NR, Mao JH and Graham GJ. LD78beta, a non-allelic variant of human MIP-1alpha (LD78alpha), has enhanced receptor interactions and potent HIV suppressive activity. *J Biol Chem* 1999;274:17478-83

141. Aquaro S, Menten P, Struyf S, et al. The LD78beta isoform of MIP-1alpha is the most potent CC-chemokine in inhibiting CCR5-dependent human immunodeficiency virus type 1 replication in human macrophages. *J Virol* 2001;75:4402-6
142. Kawase T, Akatsuka Y, Torikai H, et al. Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. *Blood* 2007;110:1055-63
143. Bonnevie-Nielsen V, Field LL, Lu S, et al. Variation in antiviral 2',5'-oligoadenylate synthetase (2'5'AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. *Am J Hum Genet* 2005;76:623-33
144. Valentonyte R, Hampe J, Huse K, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* 2005;37:357-64
145. Roca X, Sachidanandam R and Krainer AR. Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res* 2003;31:6321-33
146. Harris NL, Senapathy P. Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res* 1990;18:3015-9
147. Zhang L, Luo L. Splice site prediction with quadratic discriminant analysis using diversity measure. *Nucleic Acids Res* 2003;31:6214-20
148. Liu L, Ho YK and Yau S. Prediction of primate splice site using inhomogeneous Markov chain and neural network. *DNA Cell Biol* 2007;26:477-83
149. Kashiwabara AY, Vieira DC, Machado-Lima A and Durham AM. Splice site prediction using stochastic regular grammars. *Genet Mol Res* 2007;6:105-15
150. Lu J, Honczarenko M and Sloan SR. Independent expression of the two paralogous CCL4 genes in monocytes and B lymphocytes. *Immunogenetics* 2004;55:706-11
151. Ali S, Palmer AC, Banerjee B, Fritchley SJ and Kirby JA. Examination of the function of RANTES, MIP-1alpha, and MIP-1beta following interaction with heparin-like glycosaminoglycans. *J Biol Chem* 2000;275:11721-7

152. McCornack MA, Boren DM and LiWang PJ. Glycosaminoglycan disaccharide alters the dimer dissociation constant of the chemokine MIP-1 beta. *Biochemistry* 2004;43:10090-101
153. Laurence JS, Blanpain C, De Leener A, Parmentier M and LiWang PJ. Importance of basic residues and quaternary structure in the function of MIP-1 beta: CCR5 binding and cell surface sugar interactions. *Biochemistry* 2001;40:4990-9
154. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 2005;76:8-32
155. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007;315:848-53
156. White SJ, Breuning MH and den Dunnen JT. Detecting copy number changes in genomic DNA: MAPH and MLPA. *Methods Cell Biol* 2004;75:751-68
157. Sellner LN, Taylor GR. MLPA and MAPH: new techniques for detection of gene deletions. *Hum Mutat* 2004;23:413-9
158. Aitman TJ, Dong R, Vyse TJ, et al. Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 2006;439:851-5
159. Linzmeier RM, Ganz T. Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics* 2005;86:423-30
160. Fellermann K, Stange DE, Schaeffeler E, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 2006;79:439-48