DISCOVERY AND CHARACTERISATION OF ROBUST CHEMOKINE SIGNALLING NETWORKS IN INFLAMMATORY DISEASE USING SINGLE-CELL RNA SEQUENCING AND PROTEIN STRUCTURE PREDICTION

Charlie J. Clark

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

DPhil in Genomic Medicine and Statistics

2026

Acknowledgments

I am deeply grateful to Shoumo Bhattacharya for his supervision, scientific judgement, and sustained support throughout this work, and to Garrett Morris for his guidance, insight, and encouragement at key stages of the project. I am especially grateful to Serena Vales for her friendship, generosity, and unwavering support throughout the highs and lows of this work. I thank my labmates Megan Payne, Gintare Smagurauskaite, Soumyanetra Chandra, Stepan Denisov, Emma Wilde, Shirley Nannono-Senabulya, Philomena Mburu, and Graham Davies for their generosity with ideas, discussion, and day-to-day support in the lab.

Contents

Acknowledgments **i**

Abstract **xii**

Declaration **xiv**

List of Abbreviations **xv**

Glossary **xvii**

1 Introduction **2**

1.1 Pharmacological strategies for complex disease 2

1.2 Immune-mediated inflammatory diseases 3

1.2.1 The twelve diseases under study 4

1.3 Cytokine signalling in inflammation 8

1.3.1 The NF-κB axis 10

1.3.2 The STAT axis 11

1.4 The chemokine system 16

1.4.1 Structure and classification 16

1.4.2 Structural basis of chemokine-receptor interactions 17

1.4.3 Chemokine receptor promiscuity and ligand degeneracy 19

1.4.4 Atypical chemokine receptors 21

1.4.5 Chemokine gradients and glycosaminoglycan interactions . . . 21

1.4.6 Evolution of the chemokine system 22

1.4.7 Chemokines in immune-mediated inflammatory disease 24
1.5 Therapeutic targeting of cytokine networks 25
1.5.1 NF-κB and STAT pathway inhibitors 25
1.5.2 Failure of chemokine-targeted therapies 26
1.6 System complexity and network robustness 28
1.6.1 Complexity in the chemokine system 28
1.6.2 Biological robustness 29
1.6.3 The redundancy debate 30
1.7 Network biology and graph-theoretic analysis 31
1.7.1 Bipartite networks 31
1.7.2 Measuring network redundancy 32
1.7.3 Measuring network robustness 32
1.7.4 Prior network-level analyses of chemokine signalling 33
1.7.5 Application to cytokine signalling 34
1.8 Tick evasins as multi-chemokine inhibitors 35
1.8.1 Multi-target therapeutic approaches 36
1.9 Single-cell RNA sequencing in immunology 38
1.10 Computational structural biology and AlphaFold 40
1.11 Thesis aims and objectives 42
1.12 Thesis organisation 44

2 Materials and Methods **46**

2.1 Public databases 46
2.1.1 Gene Expression Omnibus 46
2.1.2 CELLxGENE Discover 46
2.1.3 GENCODE 47
2.1.4 Ensembl 47

2.1.5 UniProt 47
2.1.6 Gene Ontology 47
2.1.7 Guide to Pharmacology (IUPHAR/BPS) 48
2.1.8 BioGRID 48
2.1.9 IntAct 48
2.1.10 IMEx 48
2.1.11 Molecular INTeraction database (MINT) 49
2.1.12 PSICQUIC interaction providers 49
2.1.13 Protein Data Bank 49
2.1.14 AlphaFold 3 structure prediction server 49
2.2 Single-cell transcriptomic data sources 50
2.3 Data processing pipeline 52
2.3.1 Data wrangling and annotation 54
2.3.2 Quality control 54
2.3.3 Normalisation 55
2.4 Cell type classification 55
2.5 Identification of cytokine and receptor genes 56
2.6 Disease-level expression quantification 57
2.7 Network construction 57
2.7.1 Ligand-receptor interaction data 57
2.7.2 Network representation 58
2.8 Network redundancy analysis 60
2.9 Network robustness analysis 61
2.9.1 Secondary extinction analysis 61
2.9.2 Largest connected component robustness analysis 62
2.10 Identification of chemokine interactions 63

2.11 Structural prediction with AlphaFold 3 64
2.12 Structural analysis methods 65
2.12.1 RMSD calculation 65
2.12.2 Analysis of dimeric partner binding sites on chemokines 65
2.12.3 Binding site overlap analysis 65
2.12.4 Analysis of chemokine binding sites on evasins 66
2.12.5 Conformational plasticity analysis 66
2.12.6 Interface residue conservation 67
2.13 Statistical analysis 67
2.14 Software used 67

3 Analysis of Single-Cell RNA-Sequencing Data from Twelve Immunoinflammatory Diseases 68

3.1 Introduction 68
3.2 Results 70
3.2.1 ScRNA-seq data from twelve immune-mediated inflammatory diseases were selected for analysis 70
3.2.2 A computational pipeline was constructed to process these datasets uniformly 71
3.2.3 CellTypist classified cells as immune or non-immune 72
3.2.4 Immune cell distribution was visualised in heatmaps and UMAP 73
3.2.5 Fold change in cytokine and receptor expression was evaluated for nine diseases 77
3.2.6 Chemokine ligand-receptor asymmetry was conserved across diseases 81
3.2.7 Spatial single-cell RNA-seq datasets 82
3.3 Discussion 87

4 Construction and Analysis of Chemokine Signalling Networks in Inflammatory Disease 91

4.1 Introduction . . . 91

4.2 Results . . . 94

4.2.1 Construction of disease-specific cytokine networks . . . 94

4.2.2 Quantification of network edge redundancy . . . 96

4.2.3 Promiscuity drives network redundancy . . . 98

4.2.4 Functional robustness of disease networks . . . 101

4.2.5 Structural robustness of disease networks . . . 105

4.2.6 Sensitivity analysis of expression thresholds . . . 107

4.2.7 Spatial single-cell RNA-seq network analysis in psoriasis and ulcerative colitis . . . 111

4.3 Discussion . . . 117

5 Characterisation and Structural Analysis of the Ability of Tick Evasins to Attack Chemokine Networks in Inflammatory Disease 124

5.1 Introduction . . . 124

5.2 Results . . . 127

5.2.1 Compilation of chemokine interaction data . . . 127

5.2.2 AlphaFold 3 structure prediction . . . 132

5.2.3 Validation against experimental structures . . . 136

5.2.4 Evasins sterically occlude functional chemokine sites . . . 138

5.2.5 Systematic analysis of binding site overlap . . . 138

5.2.6 Quantification of binding site overlap . . . 140

5.2.7 Conformational plasticity at evasin binding interfaces . . . 141

5.2.8 Conservation of interface residue identity . . . 145

5.3 Discussion . . . 147

6 Discussion **152**

6.1 Inflamed tissues show conserved immune-cell enrichment and chemokine ligand excess 152

6.2 Chemokine networks show greater redundancy and robustness than comparator cytokine networks 153

6.3 Tick evasins reveal a structural route to broad chemokine inhibition . 154

6.4 Expression, topology, and structure converge on multi-target inhibition 155

6.5 Therapeutic implications 157

6.6 Reframing the redundancy debate 160

6.7 Evolutionary convergence 162

6.8 Systemic interpretation 165

6.9 Limitations 166

6.10 Future directions 169

6.11 Concluding remarks 171

References **173**

A Disease-specific cytokine networks **200**

List of Figures

1.1 The chemokine signalling axis 9
1.2 The NF-κB signalling axis 12
1.3 The STAT signalling axis 13
1.4 Representative CC and CXC chemokine structures 18
1.5 Chemokine-receptor interaction network 20

2.1 Computational pipeline for scRNA-seq analysis 53

3.1 Immune cell marker expression heatmap 73
3.2 Cell type distribution heatmap 74
3.3 UMAP projection of immune cell populations 75
3.4 Cell type abundances 76
3.5 Fold change across CK:CKR, NF-κB, and STAT axes 78
3.6 Transcripts per 10k for CK:CKR, NF-κB, and STAT 80
3.7 Spatial scRNA-seq panel coverage 83
3.8 Spatial scRNA-seq maps of cells in tissue 84
3.9 Spatial scRNA-seq cell-type composition 85
3.10 Spatial scRNA-seq CK:CKR, NF-κB, and STAT expression 86

4.1 Introduction to chord diagram visualisation 95
4.2 Redundancy index across signalling axes 99
4.3 Redundancy and promiscuity 100
4.4 Sequential ligand removal in the SLE NF-κB network 101
4.5 Secondary extinction curves for Crohn's disease 103

4.6 Secondary extinction robustness across networks 104
4.7 Largest connected component robustness curves for Crohn's disease . 106
4.8 Largest connected component robustness across networks 108
4.9 Sensitivity analysis of expression thresholds 110
4.10 Spatial scRNA-seq tile-level cytokine networks at 1500 micrometres . 112
4.11 Spatial scRNA-seq tile-level redundancy and robustness distributions 114
4.12 Spatial scRNA-seq tile-size and expression-threshold sensitivity 116
4.13 Spatial scRNA-seq pairwise tests across tile-level networks 117

5.1 Pipeline for structural prediction of chemokine network interactions . 127
5.2 Chord diagram of chemokine:evasin interactions 129
5.3 Chord diagram of chemokine:chemokine receptor interactions 130
5.4 Chord diagram of chemokine:chemokine interactions 131
5.5 Example AlphaFold 3 models of class A and class B evasin complexes 133
5.6 Example AlphaFold 3 model of a chemokine:chemokine receptor complex 134
5.7 Example AlphaFold 3 model of a chemokine:chemokine complex . . . 135
5.8 RMSD between AlphaFold 3 models and PDB structures 137
5.9 Multiple sequence alignment of chemokines used for contact-profile analysis 139
5.10 Binding site contact frequency across chemokine interactions 140
5.11 Pipeline for evasin interface analysis 142
5.12 Evasin interface overlay across chemokine partners 143
5.13 EVA1 interface and non-interface TM-score matrices 144
5.14 Interface and non-interface TM-score distributions 145
5.15 Conservation of evasin interface residue identity 146

A.1 Atherosclerosis ligand-receptor networks 201
A.2 Crohn's disease ligand-receptor networks 202

A.3 Discoid lupus erythematosus ligand-receptor networks 203
A.4 Endometriosis ligand-receptor networks 204
A.5 Idiopathic pulmonary fibrosis ligand-receptor networks 205
A.6 Psoriasis ligand-receptor networks 206
A.7 Psoriatic arthritis ligand-receptor networks 207
A.8 Rheumatoid arthritis ligand-receptor networks 208
A.9 Sjögren's syndrome ligand-receptor networks 209
A.10 Systemic lupus erythematosus ligand-receptor networks 210
A.11 Type 1 diabetes ligand-receptor networks 211
A.12 Ulcerative colitis ligand-receptor networks 212

List of Tables

1.1 The twelve immune-mediated inflammatory diseases 5

1.2 Ligand and receptor entries 14

1.3 Chemokine receptor antagonists evaluated in clinical trials 26

1.4 FDA-approved drugs targeting chemokine receptors 27

2.1 Dataset inclusion criteria 50

Abstract

Immune-mediated inflammatory diseases affect hundreds of millions of people worldwide, yet therapies targeting chemokines, mediators of immune cell recruitment, have consistently failed in clinical trials despite three decades of effort to develop them. In contrast, drugs targeting the NF-κB and STAT cytokine pathways have achieved clinical success. This thesis tests the hypothesis that chemokine-receptor networks possess redundancy and robustness that exceeds that of other cytokine networks, providing a mechanistic explanation for this therapeutic disparity.

Single-cell RNA-sequencing data were analysed from twelve immune-mediated inflammatory diseases spanning distinct organ systems: atherosclerosis, Crohn's disease, discoid lupus erythematosus, endometriosis, idiopathic pulmonary fibrosis, psoriasis, psoriatic arthritis, rheumatoid arthritis, Sjögren's syndrome, systemic lupus erythematosus, type 1 diabetes, and ulcerative colitis. A computational pipeline was created and used to characterise cytokine expression patterns, and 36 ligand-receptor interaction networks were constructed across twelve diseases and three signalling mechanisms.

Chemokine-receptor networks displayed greater edge redundancy than NF-κB or STAT networks. This redundancy correlated strongly with ligand-receptor promiscuity, particularly receptor-side promiscuity. Robustness analyses showed that chemokine networks maintained receptor function and structural integrity longer under simulated perturbation than other network types.

In two public spatial scRNA-seq datasets from psoriasis lesional skin and ulcerative

colitis inflamed rectal tissue, the spatial scRNA-seq analysis showed the same pattern: CK:CKR networks were more redundant and robust than NF-κB and STAT networks. This agreement showed that the network pattern persisted when ligand-receptor edges were restricted using measured cell positions.

To investigate how natural inhibitors might overcome chemokine network robustness, AlphaFold 3 was used to model 523 chemokine interactions including those with tick evasins, proteins that bind multiple chemokines simultaneously. Evasins were predicted to achieve broad-spectrum binding through conformational variability within substantially overlapping binding surfaces, while permitting some partner-specific positioning. Evasin binding was predicted to function by sterically occluding chemokine surfaces required for receptor interaction and dimerisation.

Overall, this work provides evidence that chemokine network redundancy may contribute to the failure of single-target inhibitors and suggests that multi-chemokine inhibitors mimicking evasin binding mechanisms may provide effective therapeutic strategies for immune-mediated inflammatory diseases. Direct evidence for or against this network-level explanation had not previously been gathered across diseases and cytokine signalling axes.

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution. Generative AI use was limited to grammar checking, formatting, and code generation. ChatGPT 5.4, provided by OpenAI, was used for grammar checking and formatting only. Opus 4.6, provided by Anthropic, was used for code generation. All generative-AI-assisted code was reviewed by me before use. I declare that this use complied with University, divisional, and departmental guidance on generative AI. Generative AI was not used to produce substantive original writing, analyse data, generate plots or data visualisations directly from prompts, or handle private or confidential data.

Charlie J. Clark

2026

List of Abbreviations

ACKR	Atypical chemokine receptor
CCL	CC chemokine ligand
CCR	CC chemokine receptor
CKR	Chemokine receptor
CK:CKR	Chemokine:chemokine receptor axis
CRS	Chemokine recognition site
CXCL	CXC chemokine ligand
CXCR	CXC chemokine receptor
DLE	Discoid lupus erythematosus
GPCR	G protein-coupled receptor
IBD	Inflammatory bowel disease
IFN	Interferon
IL	Interleukin
IMID	Immune-mediated inflammatory disease
IPF	Idiopathic pulmonary fibrosis
JAK	Janus kinase
LCC	Largest connected component
MSA	Multiple sequence alignment
NF-κB	Nuclear factor κ-light-chain-enhancer of activated B cells
PAE	Predicted aligned error
pLDDT	Predicted local distance difference test
PsA	Psoriatic arthritis
RA	Rheumatoid arthritis
RMSD	Root mean square deviation

scRNA-seq	Single-cell RNA sequencing
spatial scRNA-seq	Spatial single-cell RNA sequencing
SDC	Sørensen-Dice coefficient
SLE	Systemic lupus erythematosus
STAT	Signal transducer and activator of transcription
TM-score	Template modelling score
TNF	Tumour necrosis factor
TNFR	Tumour necrosis factor receptor
TP10K	Transcripts per 10,000
UMAP	Uniform Manifold Approximation and Projection

Glossary

Biased agonism. *The phenomenon whereby different ligands acting at the same receptor produce qualitatively different signalling outcomes, activating distinct downstream pathways.*

Bipartite network. *A network in which nodes are divided into two disjoint sets, with edges connecting only nodes from different sets. Ligand-receptor networks are bipartite, with ligands in one set and receptors in the other.*

Chemokine. *A family of small cytokines that direct immune cell migration through concentration gradients. Named for their chemoattractant properties; classified into CC, CXC, CX3C, and XC subfamilies based on N-terminal cysteine arrangement.*

CK:CKR axis. *The chemokine:chemokine receptor signalling axis, comprising chemokine ligands and the chemokine receptors that mediate leukocyte recruitment. CKR alone denotes chemokine receptors.*

Conformational plasticity. *The ability of a protein binding interface to adopt multiple conformations, enabling binding to structurally diverse partners.*

Edge redundancy. *A network property quantifying the excess of edges beyond the minimum required to maintain connectivity. Calculated as* $R = 1 - |E_{\min}|/|E|$*, where* $|E|$ *is total edges and* $|E_{\min}|$ *is the minimum edge cover.*

Evasin. *A family of chemokine-binding proteins secreted in tick saliva that inhibit multiple chemokines simultaneously. Class A evasins bind CC chemokines; class B evasins bind CXC chemokines.*

Knottin scaffold. *A protein fold characterised by three disulfide bonds forming a knot-like topology, providing high stability. Class B evasins adopt this fold.*

Largest connected component. *The largest subset of network nodes in which any two nodes are connected by a path. Used to measure network fragmentation under perturbation.*

Minimum edge cover. *The smallest set of edges that includes every node in a network at least once. Represents the minimal connectivity required to reach all nodes.*

Primary extinction. *In network robustness analysis, the deliberate removal of a node and all its edges, simulating loss of a signalling component.*

Promiscuity. *In chemokine biology, the ability of a ligand to bind multiple receptors or a receptor to bind multiple ligands. Quantified as the average number of binding partners per molecule.*

Robustness. *The capacity of a system to maintain function despite perturbation. In network terms, measured by resistance to node or edge removal.*

Secondary extinction. *In network robustness analysis, the loss of a receptor node that occurs when all its ligand partners have been removed through primary extinctions.*

Spatial scRNA-seq. *Single-cell RNA sequencing in which gene expression is measured for individual cells while retaining each cell's position in the tissue.*

Steric occlusion. *Inhibition of protein-protein interaction by physical blockade of the binding interface, preventing the native partner from accessing its binding site.*

Chapter 1

Introduction

1.1 Pharmacological strategies for complex disease

A longstanding strategy in modern pharmacology is selective intervention: a therapeutic should act on a defined disease-causing target while sparing other biological processes[1]. Paul Ehrlich's early twentieth-century "magic bullet" concept formalised this aim in antimicrobial chemotherapy and later influenced target-based drug discovery[1]. This approach is most likely to succeed when disease depends on a non-redundant molecular vulnerability and inhibition of that vulnerability is sufficient to alter the clinical phenotype[2].

However, complex diseases may lack a single such dependency because biological networks may contain overlapping interactions and compensatory pathways that can preserve function after one component is inhibited[2–4]. This observation motivated rational multi-target pharmacology. Roth et al. used the term "magic shotgun" for compounds designed to act on a selected set of targets rather than one target alone[5]. In this usage, multi-target activity is intentional and bounded; it is distinct from uncharacterised off-target activity[5].

Multi-target effects can be achieved through drug combinations or single compounds with defined polypharmacology[2,5]. Biological agents that bind more than one related

ligand provide another route to multi-target activity[6]. Network pharmacology provides a framework for selecting such interventions by relating drug action to the organisation and robustness of the disease-associated network[2]. The appropriate strategy therefore depends on system architecture: a selective inhibitor may suffice when its target is a bottleneck, whereas a redundant network may require simultaneous perturbation of two or more components[2–4].

This distinction frames the central problem examined in this thesis. In immune-mediated inflammatory diseases, drugs directed at individual chemokines or chemokine receptors have repeatedly failed in inflammatory indications, while selective interventions within the NF-κB and STAT axes have achieved clinical efficacy[7–10]. This thesis tests whether differences in ligand-receptor network redundancy and robustness help explain that contrast, and whether broad chemokine binding by tick evasins provides a structural model for rational multi-target inhibition[6].

1.2 Immune-mediated inflammatory diseases

Immune-mediated inflammatory diseases (IMIDs) are a heterogeneous group of chronic conditions characterised by dysregulated immune responses that cause persistent tissue inflammation and damage[11]. These diseases collectively affect hundreds of millions of people worldwide and include rheumatoid arthritis, inflammatory bowel disease, psoriasis, systemic lupus erythematosus, type 1 diabetes, and atherosclerosis[12,13]. Despite diverse clinical manifestations and affected organ systems, IMIDs share common pathological mechanisms, including the mechanism of cytokine signalling perpetuating chronic inflammation, driving immune cell infiltration into target tissues and causing progressive tissue damage[11,14].

The global burden of IMIDs is substantial and growing. In 2020, the age-standardised

global prevalence of rheumatoid arthritis was 208.8 cases per 100 000 population[15]. Inflammatory bowel disease prevalence continues to rise, particularly in newly industrialised nations[16]. Atherosclerosis, an inflammatory disease of the arterial wall, is the leading cause of cardiovascular mortality worldwide[17,18]. An estimated 4.62 million new cases of psoriasis occurred globally in 2019[13]. Existing therapies, including TNF blockers, IL-6 receptor antagonists, and JAK inhibitors, have improved outcomes for many patients, but primary non-response, secondary loss of response, and long-term adverse effects remain significant clinical problems[9,10,19,20].

The shared pathological features of IMIDs justify their study as a group rather than as isolated conditions[11]. Despite affecting different organ systems, from the arterial wall in atherosclerosis to the synovium in rheumatoid arthritis to the colonic mucosa in inflammatory bowel disease, IMIDs are united by common immunological mechanisms, such as T-cell and macrophage infiltration, pro-inflammatory cytokine production, and chemokine-directed leukocyte recruitment[11,14]. The success of therapies that target shared cytokine pathways across multiple IMIDs, such as TNF inhibitors in both rheumatoid arthritis and Crohn's disease, confirms that common mechanisms operate across these conditions[9,10]. Based on this observation, this thesis aims to exploit this commonality by analysing cytokine networks across twelve IMIDs simultaneously, testing whether shared network properties explain the consistent pattern of chemokine therapeutic failure observed across diseases.

1.2.1 The twelve diseases under study

Here, cytokine signalling networks are examined across twelve IMIDs selected because they span distinct organ systems, contribute substantially to global disease burden, and have publicly available single-cell RNA-sequencing datasets. The focus on publicly available single-cell RNA-sequencing datasets reflects the scope of the project, since

manual collection of human samples from multiple diseases would exceed the scope of a single DPhil thesis. **Table 1.1** provides a comparative overview of these conditions.

Table 1.1: Comparative overview of the twelve immune-mediated inflammatory diseases examined in this thesis. Prevalence is given per 100,000 population unless otherwise stated. Approved therapies refer to drugs that directly target an NF-κB- or JAK–STAT-linked cytokine, receptor, or JAK, with regulatory approval for each indication.

Disease	Tissue	Prevalence	Approved therapy
Atherosclerosis	Arterial wall/atherosclerotic plaque[18,21]	815.76 million adults aged 30–79 with carotid plaque worldwide in 2020[22]	None identified[23,24]
Type 1 diabetes	Pancreatic islets[25]	8.4M globally[26]	None[27]
Crohn's disease	GI tract[28]	84.2[29]	Infliximab, ustekinumab[28]
Ulcerative colitis	Colonic mucosa[30]	120.4[29]	Infliximab, tofacitinib[30]
IPF	Lung[31]	3.3–45.1 across studied countries[32]	None[31]
Psoriasis	Skin[33]	503.6[34]	Secukinumab[35], guselkumab[36]
SLE	Multiple organs[37]	44[38]	Belimumab, anifrolumab[37]
DLE	Skin[39]	6.5, age-adjusted Manhattan estimate[40]	None[41]
Sjögren's syndrome	Salivary glands[42]	61[43]	None[44]
Rheumatoid arthritis	Synovium[45]	209[15]	Tocilizumab, tofacitinib[10]
Psoriatic arthritis	Synovium[46]	133[47]	Secukinumab, adalimumab[48]
Endometriosis	Ectopic endometrial lesions, including peritoneal, ovarian, and fallopian tube sites[49,50]	Not reported	None[49]

Atherosclerosis is a chronic inflammatory disease of the arterial wall characterised

by lipid accumulation, immune cell infiltration, and progressive plaque formation[18,21]. It underlies most cardiovascular events, including myocardial infarction and stroke, making it the leading cause of mortality worldwide[17].

Type 1 diabetes results from autoimmune destruction of insulin-producing β-cells in the pancreatic islets[25,27]. Chemokines direct autoreactive T-cells to the islets, where they mediate β-cell killing and progressive loss of insulin secretion[25,27].

Crohn's disease and ulcerative colitis are the two major forms of inflammatory bowel disease[28,30,51]. Crohn's disease can affect any part of the gastrointestinal tract with transmural inflammation, while ulcerative colitis is confined to the colonic mucosa[28,30]. Both involve dysregulated mucosal immunity and chemokine-driven leukocyte infiltration[51,52].

Idiopathic pulmonary fibrosis is a progressive fibrotic lung disease characterised by aberrant wound healing, fibroblast activation, and extracellular matrix deposition[31]. Inflammatory chemokines contribute to disease progression through recruitment of profibrotic immune cells[31].

Psoriasis is a chronic inflammatory skin disease characterised by keratinocyte hyperproliferation and immune cell infiltration[33,35]. The IL-23/IL-17 axis is involved, and chemokines direct T-cell and neutrophil recruitment to the epidermis[33,35].

Systemic lupus erythematosus and discoid lupus erythematosus are systemic and cutaneous forms of lupus, respectively[37,39]. SLE is characterised by autoantibody production and immune complex deposition affecting multiple organs, while DLE primarily affects the skin[37,53].

Sjögren's syndrome is an autoimmune disease affecting exocrine glands, especially the salivary and lacrimal glands, leading to dry mouth and dry eyes[42,44,54]. Chemokines are involved in the ectopic lymphoid structures that form within affected glands[42,54].

Rheumatoid arthritis and psoriatic arthritis are inflammatory arthropathies characterised by synovial inflammation and structural joint damage[46]. In RA, multiple CC and CXC chemokines are expressed in joint tissues[45,55].

Endometriosis is characterised by the presence of endometrial tissue outside the uterus, including in the peritoneal cavity[49,50]. Mesothelial cells adjacent to endometriomas overexpress inflammatory mediators, including MIF[50]. Pelvic pain is reported by 90% of people with endometriosis[49].

These twelve conditions can be grouped by their dominant immunological mechanisms. Rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes, and Sjögren's syndrome are primarily autoimmune, driven by adaptive immune responses against self-antigens[27,37,44,56]. Psoriasis and inflammatory bowel disease involve both innate and adaptive immune dysregulation, with the IL-23/IL-17 axis central to psoriasis and mucosal immune dysregulation central to IBD[28,33,35,52]. Idiopathic pulmonary fibrosis and endometriosis involve inflammatory fibrotic processes where immune cell recruitment contributes to tissue remodelling[31,50]. Atherosclerosis is a vascular inflammatory disease in which lipid-driven immune responses in the arterial wall drive plaque formation and can contribute to plaque rupture[17,18,21]. Despite these mechanistic differences, chemokines are implicated in leukocyte recruitment in all twelve conditions[57,58]. The diversity of the selection is important because conservation of chemokine network properties across diseases with different aetiologies, organ systems, and immune mechanisms would indicate that network redundancy is a general feature of chemokine signalling rather than a property of any specific disease context.

1.3 Cytokine signalling in inflammation

Cytokines orchestrate immune and inflammatory responses through a limited set of downstream signalling mechanisms that result in immune-mediated inflammatory disease[11,14]. Broadly, these fall into three axes[11,14]. Chemokines signal via G protein-coupled chemokine receptors (CKRs) to direct immune cell trafficking[57,58]. This thesis refers to the chemokine:chemokine receptor axis as the CK:CKR axis, while CKR alone denotes chemokine receptors. Members of the tumour necrosis factor (TNF), interleukin-1 (IL-1), and IL-17 families converge on the transcription factor NF-κB[14,59]. Interferons, together with the IL-2, IL-3, IL-6, IL-10, and IL-12 families, predominantly activate Signal Transducer and Activator of Transcription (STAT) transcription factors[14,60].

The organisation of the chemokine signalling axis is shown in **Figure 1.1**. It comprises chemokine ligands, their receptors, and the immune cells that express those receptors.

Although distinct, these pathways are interconnected: NF-κB and STAT signals intersect via promoter co-occupancy, physical interactions, and co-activator sharing, integrating diverse cytokine inputs into common transcriptional programs that govern inflammation and immune-cell differentiation[62]. Moreover, NF-κB and STAT pathways interface with CK:CKR signalling by directly inducing chemokine gene transcription[63,64]. TNF stimulation activates NF-κB, which binds promoter elements in CCL2, CCL5, CXCL8, and other chemokine genes, driving their transcription[59,63]. IL-6 signalling through STAT3 regulates CXC chemokines including CXCL1 in stromal cells[64]. Interferon-γ signalling through STAT1 induces the CXCR3 ligands CXCL9, CXCL10, and CXCL11[60,63].

Chemokine-driven recruitment of cytokine-producing cells then amplifies local cy-

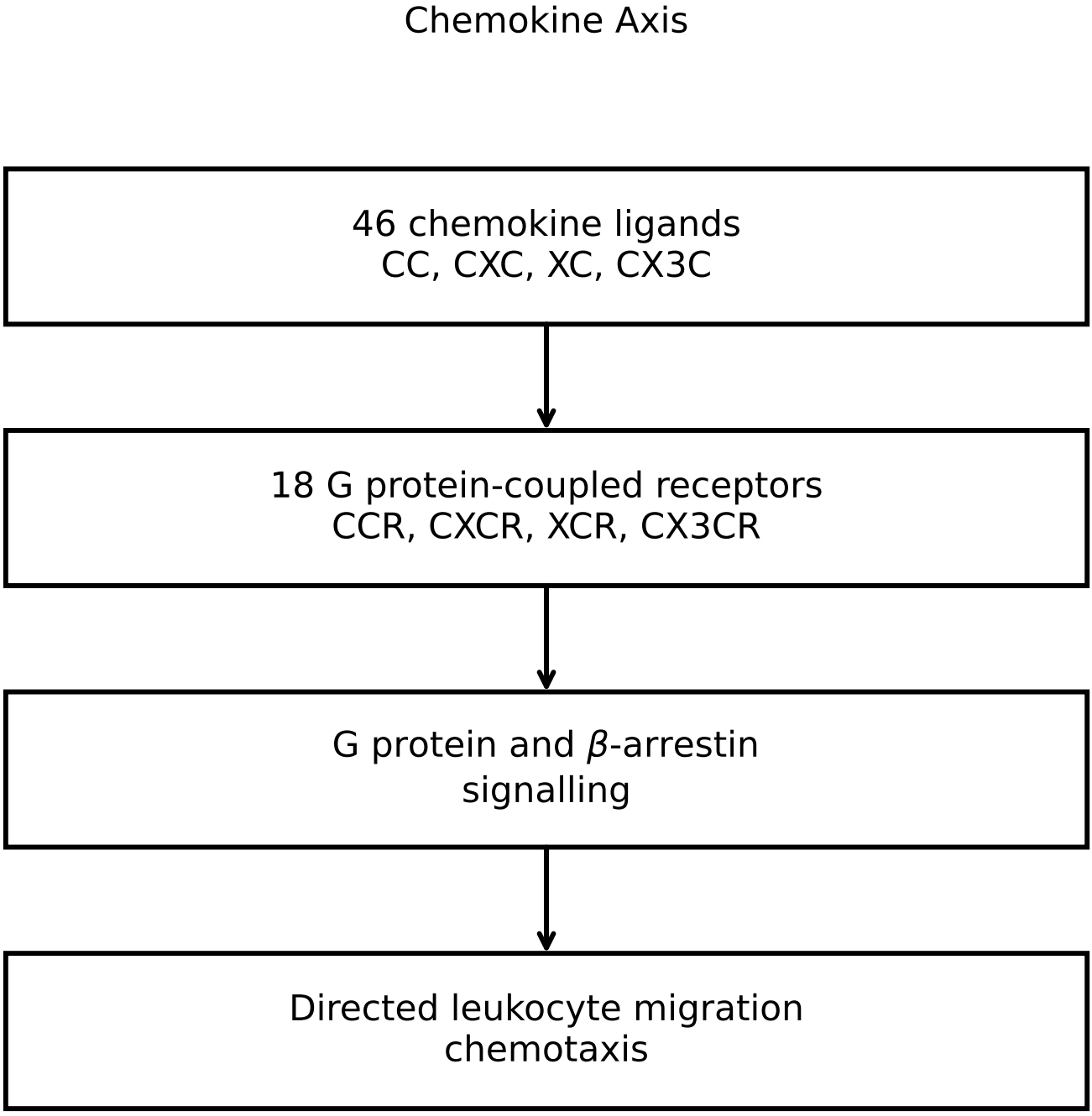

Figure 1.1: The chemokine signalling axis directs immune-cell migration through 46 chemokine ligands and 18 G protein-coupled receptors[61].

tokine production, creating feed-forward loops that escalate inflammation and tissue injury[65]. Recruited macrophages produce TNF and IL-1β, which drive further NF-κB-dependent chemokine expression; recruited T-cells produce IFN-γ, which activates STAT1-dependent chemokine transcription[14,65]. This hierarchical relationship between the signalling axes has a therapeutic implication: drugs that block NF-κB or STAT signalling may achieve indirect multi-chemokine inhibition by suppressing the transcriptional processes that drive chemokine production[63]. The three-axis framework provides the basis for the comparative analysis in this thesis, as the NF-κB and STAT axes can be used as internal controls against which chemokine network properties can be measured. All three axes operate in the same disease tissues and are constructed from the same expression data, so differences in network redundancy or robustness are expected to arise from the topology of the interactions themselves rather than from differences in data quality, analytical methods, or disease context.

1.3.1 The NF-κB axis

The NF-κB axis functions as a central hub for pro-inflammatory signalling downstream of TNF, IL-1, and IL-17 family ligands[14,59]. While each family has distinct receptors and intermediates, they converge on NF-κB nuclear translocation and pro-inflammatory gene expression[59].

TNF is the prototypic member of a family comprising more than 15 ligands that signal through TNF receptor superfamily members[14,66]. The IL-1 family includes 11 members: IL-1α/IL-1β, IL-18, IL-33, and endogenous antagonists such as IL-1RN and IL-37[67]. The IL-17 family, notably IL-17A and IL-17F, induces chemokines, neutrophil recruitment, and tissue inflammation[68]. Collectively, these families drive acute and chronic inflammation through NF-κB-dependent transcriptional processes[14,59].

The NF-κB axis is therefore a useful comparator for chemokine network analysis. It is

strongly implicated in inflammation and has been successfully targeted therapeutically, but its ligands and receptors do not form the same ligand-heavy GPCR interaction system as chemokines[9,14,59]. Comparing this axis with CK:CKR signalling helps test whether chemokine therapeutic failure reflects network architecture rather than inflammatory importance alone. The organisation of the NF-κB signalling axis is shown in **Figure 1.2**.

1.3.2 The STAT axis

The STAT axis comprises interferons and members of the IL-2, IL-3, IL-6, IL-10, and IL-12 families[60,69]. These signal via receptor-associated Janus kinases (JAKs) that phosphorylate STATs, promoting dimerisation, nuclear translocation, and DNA binding[60]. Depending on context, STAT signalling can be pro- or anti-inflammatory[14,60].

IL-2 family cytokines, comprising IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21, use receptors containing the common γ-chain (IL2RG) to regulate T-cell proliferation and differentiation[70]. IL-3 family cytokines, comprising IL-3, IL-5, and CSF2, signal via receptors containing the common β-chain (CSF2RB) and are central to myelopoiesis[71]. The IL-6 family, including IL-6 and IL-11, signals through heterodimeric complexes containing IL6ST (gp130)[64].

IL-10 family members have diverse roles: IL-10 is potently anti-inflammatory, whereas IL-19, IL-20, IL-22, and IL-24 promote epithelial responses and inflammation[72]. The IL-12 family forms heterodimeric cytokines, comprising IL-12, IL-23, IL-27, and IL-35, with divergent outcomes[73,74]. IL-12 and IL-23 amplify Th1/Th17 responses, while IL-27 and IL-35 act as negative regulators[73,74].

Interferons, comprising type I IFNα/β, type II IFNγ, and type III IFNλ, are essential for antiviral defence but can drive immunopathology[75]. Together, STAT pathways

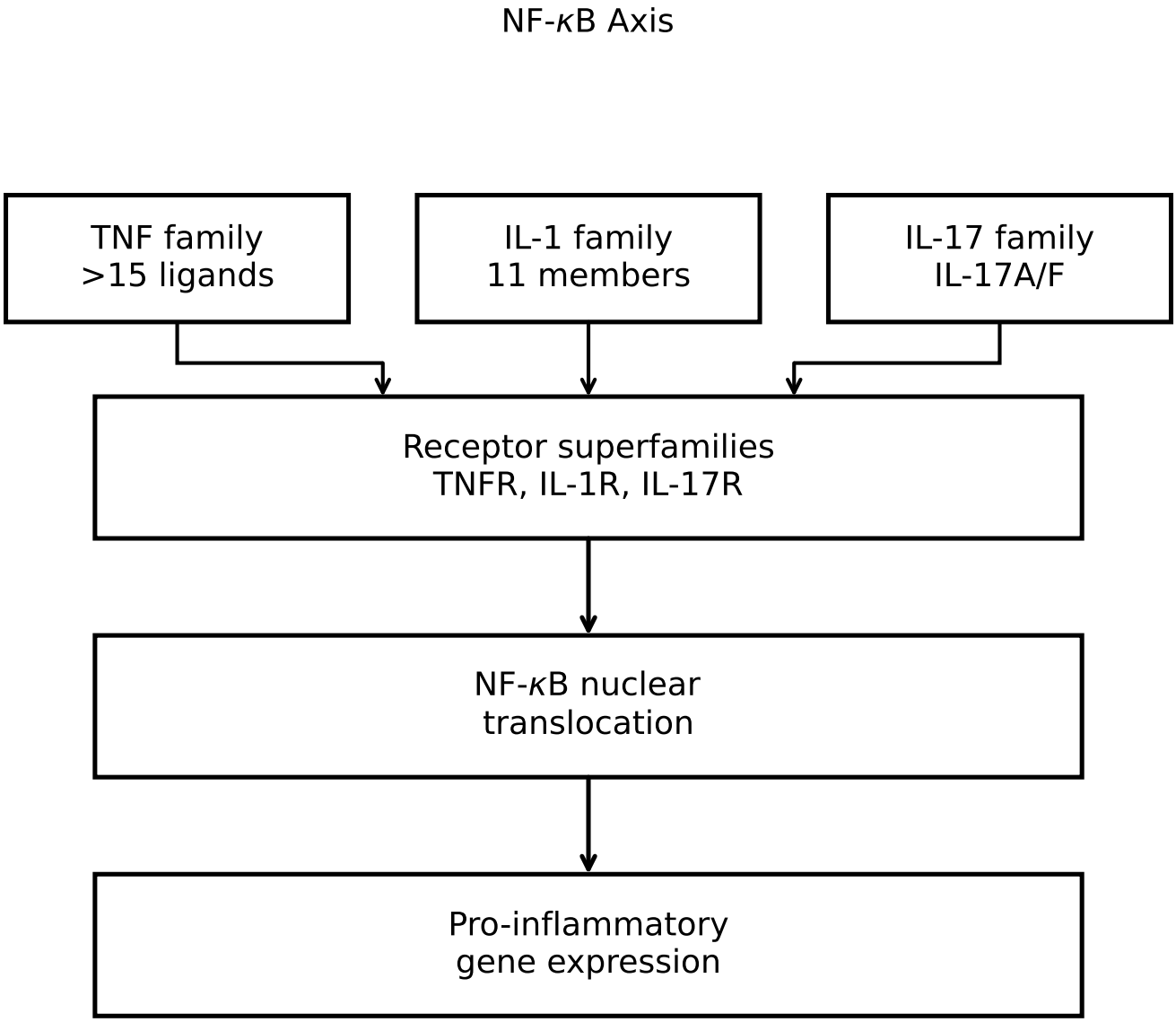

Figure 1.2: The NF-κB signalling axis drives inflammation through TNF, IL-1, and IL-17 family cytokines, which activate NF-κB nuclear translocation and pro-inflammatory gene expression[14,59].

shape both propagation and resolution of inflammation[14,60].

The STAT axis provides a second comparator because it includes clinically tractable cytokine pathways with a different receptor organisation from chemokine-GPCR signalling[9,10,60]. Including both STAT and NF-κB networks allows the chemokine axis

to be compared against two inflammatory signalling systems that operate in the same tissues but have different ligand-receptor topologies. The organisation of the STAT signalling axis is shown in **Figure 1.3**.

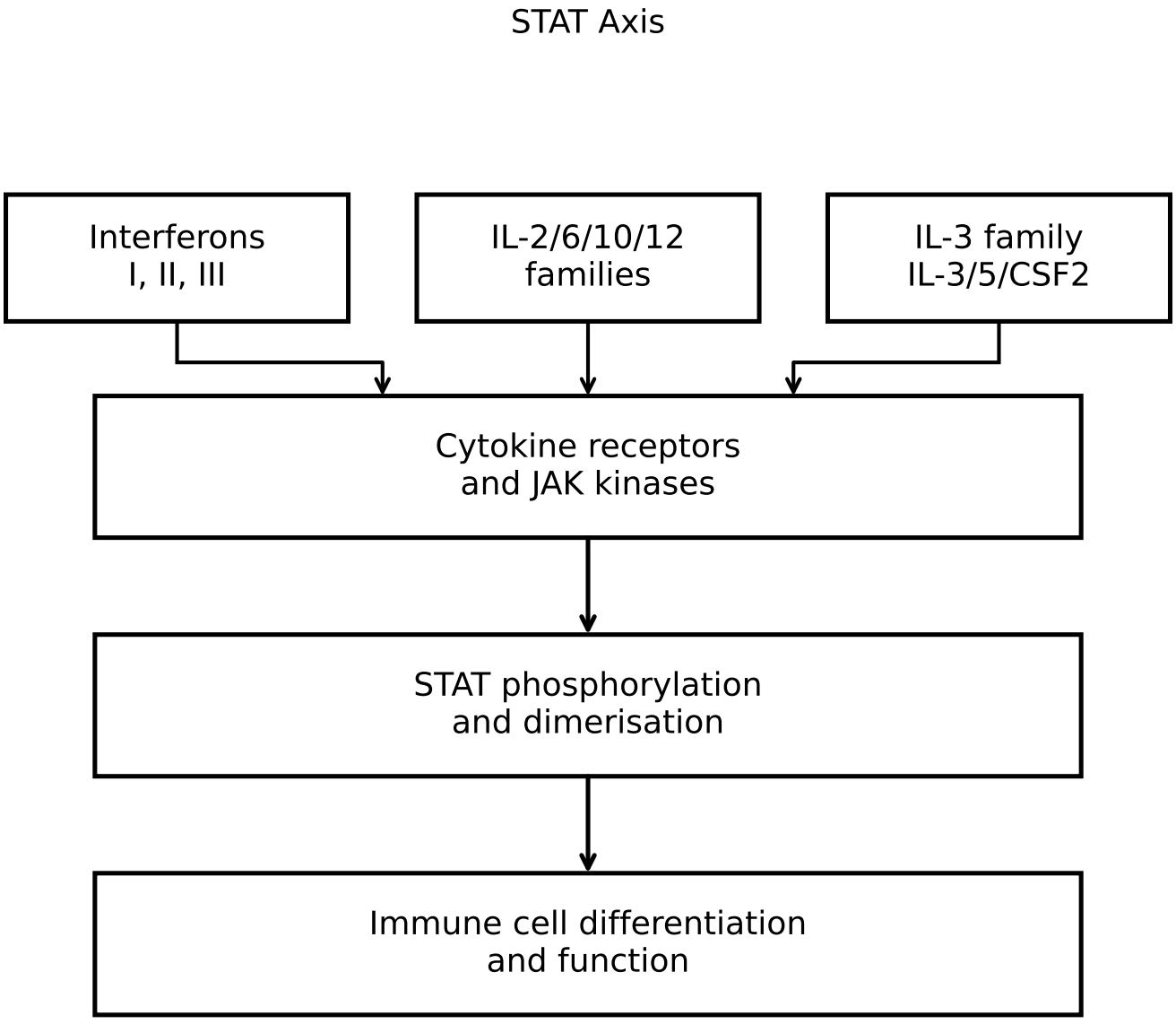

Figure 1.3: The STAT signalling axis regulates immune-cell differentiation and function through interferon and interleukin family signalling via JAK-STAT pathways[14,60].

The ligand and receptor entries used to define the three study reference networks are listed in **Table 1.2**. These lists define the repertoire from which disease-specific

networks were constructed after expression filtering [76].

Table 1.2: Ligand and receptor entries in the three study reference networks. A plus sign joins subunits and semicolons separate receptor entries.

Axis	Ligand	Receptor entries
CK:CKR	CCL1	CCR8
CK:CKR	CCL2	CCR2; CCR3; CCR5
CK:CKR	CCL3	CCR1; CCR5
CK:CKR	CCL4	CCR5
CK:CKR	CCL5	CCR1; CCR3; CCR5; CXCR3
CK:CKR	CCL7	CCR1; CCR2; CCR3
CK:CKR	CCL8	CCR1; CCR2; CCR3; CCR5
CK:CKR	CCL11	CCR3; CCR5
CK:CKR	CCL13	CCR1; CCR2; CCR3; CCR5; CXCR3
CK:CKR	CCL14	CCR1; CCR5
CK:CKR	CCL15	CCR1; CCR3
CK:CKR	CCL16	CCR1; CCR2; CCR5
CK:CKR	CCL17	CCR4
CK:CKR	CCL19	CCR7; CXCR3
CK:CKR	CCL20	CCR6; CXCR3
CK:CKR	CCL21	CCR7
CK:CKR	CCL22	CCR4
CK:CKR	CCL23	CCR1
CK:CKR	CCL24	CCR3
CK:CKR	CCL25	CCR9
CK:CKR	CCL26	CCR3
CK:CKR	CCL27	CCR10
CK:CKR	CCL28	CCR3; CCR10
CK:CKR	CX3CL1	CX3CR1
CK:CKR	CXCL1	CXCR2
CK:CKR	CXCL2	CXCR2
CK:CKR	CXCL3	CXCR2
CK:CKR	CXCL5	CXCR2
CK:CKR	CXCL6	CXCR1; CXCR2
CK:CKR	CXCL8	CXCR1; CXCR2
CK:CKR	CXCL9	CXCR3
CK:CKR	CXCL10	CXCR3
CK:CKR	CXCL11	CXCR3
CK:CKR	CXCL12	CXCR3; CXCR4
CK:CKR	CXCL13	CXCR5
CK:CKR	CXCL16	CXCR6
CK:CKR	PPBP	CXCR2
CK:CKR	XCL1	XCR1
CK:CKR	XCL2	XCR1

Table 1.2 continued

Axis	Ligand	Receptor entries
NF-κB	CD40LG	CD40
NF-κB	CD70	CD27
NF-κB	EDA	EDA2R; EDAR
NF-κB	FASLG	FAS
NF-κB	IL1A	IL1RAP + IL1R1
NF-κB	IL1B	IL1RAP + IL1R1
NF-κB	IL17A	IL17RC + IL17RA
NF-κB	IL17A + IL17F	IL17RC + IL17RA
NF-κB	IL17B	IL17RA + IL17RB
NF-κB	IL17C	IL17RE + IL17RA
NF-κB	IL17F	IL17RC + IL17RA
NF-κB	IL18	IL18R1 + IL18RAP
NF-κB	IL25	IL17RA + IL17RB
NF-κB	IL33	IL1RAP + IL1RL1
NF-κB	IL36A	IL1RAP + IL1RL2
NF-κB	IL36B	IL1RAP + IL1RL2
NF-κB	IL36G	IL1RAP + IL1RL2
NF-κB	LTA	TNFRSF1A; TNFRSF1B; TNFRSF14
NF-κB	TNF	TNFRSF1A; TNFRSF1B
NF-κB	TNFSF4	TNFRSF4
NF-κB	TNFSF8	TNFRSF8
NF-κB	TNFSF9	TNFRSF9
NF-κB	TNFSF10	TNFRSF10A; TNFRSF10B
NF-κB	TNFSF11	TNFRSF11A
NF-κB	TNFSF12	TNFRSF12A
NF-κB	TNFSF13	TNFRSF17
NF-κB	TNFSF13B	TNFRSF13B; TNFRSF13C; TNFRSF17
NF-κB	TNFSF14	LTBR; TNFRSF6B; TNFRSF14
NF-κB	TNFSF15	TNFRSF25
NF-κB	TNFSF18	TNFRSF18
STAT	CNTF	IL6ST + CNTFR + LIFR
STAT	CSF2	CSF2RA; CSF2RB + CSF2RA
STAT	CTF1	IL6ST + LIFR
STAT	IFNA2	IFNAR1 + IFNAR2
STAT	IFNA4	IFNAR1 + IFNAR2
STAT	IFNA5	IFNAR1 + IFNAR2
STAT	IFNA6	IFNAR1 + IFNAR2
STAT	IFNA7	IFNAR1 + IFNAR2
STAT	IFNA8	IFNAR1 + IFNAR2
STAT	IFNA10	IFNAR1 + IFNAR2
STAT	IFNA14	IFNAR1 + IFNAR2
STAT	IFNA16	IFNAR1 + IFNAR2
STAT	IFNA17	IFNAR1 + IFNAR2

Table 1.2 continued

Axis	Ligand	Receptor entries
STAT	IFNA21	IFNAR1 + IFNAR2
STAT	IFNB1	IFNAR1; IFNAR1 + IFNAR2
STAT	IFNE	IFNAR1 + IFNAR2
STAT	IFNG	IFNGR1 + IFNGR2
STAT	IFNK	IFNAR1 + IFNAR2
STAT	IFNL1	IL10RB + IFNLR1
STAT	IFNL2	IL10RB + IFNLR1
STAT	IFNL3	IL10RB + IFNLR1
STAT	IFNW1	IFNAR1 + IFNAR2
STAT	IL2	IL2RG + IL2RA + IL2RB
STAT	IL3	CSF2RB + IL3RA; IL3RA
STAT	IL4	IL2RG + IL4R; IL4R
STAT	IL5	CSF2RB + IL5RA; IL5RA
STAT	IL6	IL6R; IL6ST + IL6R
STAT	IL7	IL2RG + IL7R
STAT	IL9	IL2RG + IL9R
STAT	IL11	IL6ST + IL11RA
STAT	IL12A + EBI3	IL12RB2 + IL6ST; IL12RB2 + IL27RA
STAT	IL12A + IL12B	IL12RB1 + IL12RB2
STAT	IL12B + IL23A	IL12RB1 + IL23R
STAT	IL15	IL2RG + IL2RB + IL15RA; IL15RA
STAT	IL19	IL20RB + IL20RA
STAT	IL20	IL20RB + IL20RA; IL20RB + IL22RA1
STAT	IL21	IL2RG + IL21R
STAT	IL22	IL10RB + IL22RA1
STAT	IL24	IL20RB + IL20RA; IL20RB + IL22RA1
STAT	IL26	IL20RB + IL10RA
STAT	IL31	IL31RA + OSMR
STAT	LIF	IL6ST + LIFR
STAT	OSM	IL6ST + LIFR; IL6ST + OSMR

1.4 The chemokine system

1.4.1 Structure and classification

Chemokines are a family of 46 small secreted proteins that orchestrate immune-cell migration[57,58]. They fall into CC, CXC, CX3C, and XC subfamilies based on N-terminal cysteine arrangement and signal through 18 conventional CKRs, comprising the CCR,

CXCR, CX3CR, and XCR families, while atypical chemokine receptors (ACKR1–4) act primarily as scavenger receptors that modulate chemokine availability without promoting chemotaxis[58,61]. Functionally, chemokines are categorised as homeostatic or inflammatory, though many ligands display both roles[58,77].

Chemokines share a conserved disulfide-bonded tertiary structure comprising an unstructured N-terminal domain, an N-loop, a three-stranded β-sheet linked by the 30s and 40s loops, and a C-terminal α-helix[78,79]. This conserved fold has been maintained throughout vertebrate evolution despite extensive sequence divergence[58,80]. Despite their conserved fold, chemokine sequence identity is below 30% between the four major groups and approximately 30–99% within groups[61].

Representative CC and CXC chemokines share the conserved chemokine fold[78,79]. Their shared fold and distinguishing cysteine arrangements are shown in **Figure 1.4**.

1.4.2 Structural basis of chemokine-receptor interactions

Chemokine receptors are class A G protein-coupled receptors (GPCRs) that engage chemokines through a two-step binding mechanism involving four chemokine recognition sites (CRS)[57,79]. CRS0.5 is in the distal N-terminus of the receptor and interacts with the β1-strand of the chemokine, while CRS1 is in the receptor N-terminus and interacts with the N-loop and 40s loop[78,79]. CRS1.5 is between CRS1 and CRS2 and targets the conserved disulfide bond, and CRS2 comprises the transmembrane pocket and interacts with the flexible N-terminus of the chemokine[78,79].

The two-step binding model proposes that the chemokine first docks to the receptor via CRS1, positioning the chemokine N-terminus to insert into the transmembrane pocket at CRS2 and trigger receptor activation[57,79]. The reciprocal nature of this interaction, where the structured core of the chemokine engages the unstructured

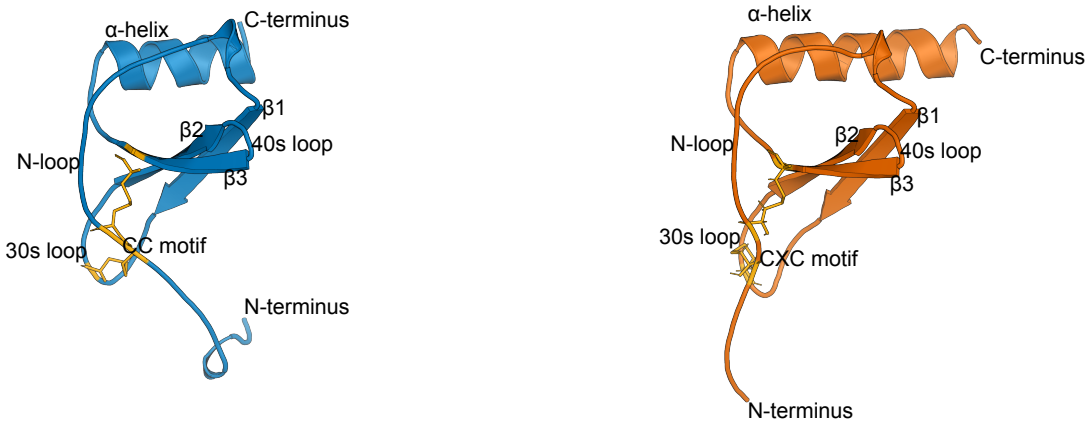

Figure 1.4: Matched-orientation cartoon representations show human CCL2 on the left in blue (CC chemokine; PDB 1DOL) and human CXCL8 on the right in orange (CXC chemokine; PDB 1IL8). Each chain displays the conserved chemokine fold, comprising the flexible N-terminus, N-loop, three-stranded β-sheet, 30s and 40s loops, and C-terminal α-helix. Gold sticks show the conserved disulfide bonds, including the N-terminal CC and CXC cysteine arrangements[78,79,81,82].

N-terminus of the receptor and vice versa, is a feature of the chemokine-GPCR system that is preserved across all experimentally characterised complexes[79,83].

Chemokines also homo- and heterodimerise, which affects receptor binding differently in CC and CXC classes[57,84]. CC-class chemokines typically cannot bind their receptors as dimers because dimerisation occludes the receptor-binding surface, whereas CXC-class chemokines can bind receptors in both monomeric and dimeric forms[57,79]. Chemokine heterodimerisation can either enhance or inhibit receptor activation: CC-type heterodimers, formed by interactions between CC motifs and N-terminal residues, enhance chemokine activity, whereas CXC-type heterodimers, formed by β1-strand interactions, inhibit function[84,85]. The functional consequences of these heteromeric

interactions add a layer of complexity beyond simple ligand-receptor binding, as the activity of a chemokine depends on which other chemokines are present in the local microenvironment[84].

1.4.3 Chemokine receptor promiscuity and ligand degeneracy

A defining feature of the chemokine system is the extensive apparent promiscuity in ligand-receptor interactions: many chemokines bind multiple receptors, and most receptors bind multiple chemokines[58,61,78]. For example, CCL5 (RANTES) activates CCR1, CCR3, and CCR5, while CCR1 responds to CCL3, CCL5, CCL7, CCL14, CCL15, CCL16, CCL23, and other ligands[61,78]. These highly interconnected interactions can cooperatively or synergistically modulate immune cell trafficking[84,85].

This "one-to-many" binding pattern creates a highly interconnected network in which individual chemokine-receptor pairs cannot be considered in isolation[61,86]. One interpretation is that promiscuity provides functional redundancy, ensuring immune cell recruitment even if individual components are lost[57,87]. An alternative interpretation emphasises context-specificity: different ligands may produce distinct signalling outcomes through the same receptor via biased agonism[79,88]. The resulting interaction pattern is shown in **Figure 1.5**.

Biased agonism refers to the ability of different ligands acting at the same receptor to preferentially activate distinct downstream pathways[79,89]. Chemokine receptors can signal through both G protein-dependent and β-arrestin-dependent pathways, and different chemokines may bias the receptor toward one pathway over the other[89]. For CXCR3, biased agonists produce different physiological outcomes: a β-arrestin-biased agonist potentiated inflammation and was more efficient at stimulating T-cell chemotaxis than a G protein-biased agonist[90]. If different chemokines binding the same receptor produce qualitatively different signalling, then apparent promiscuity

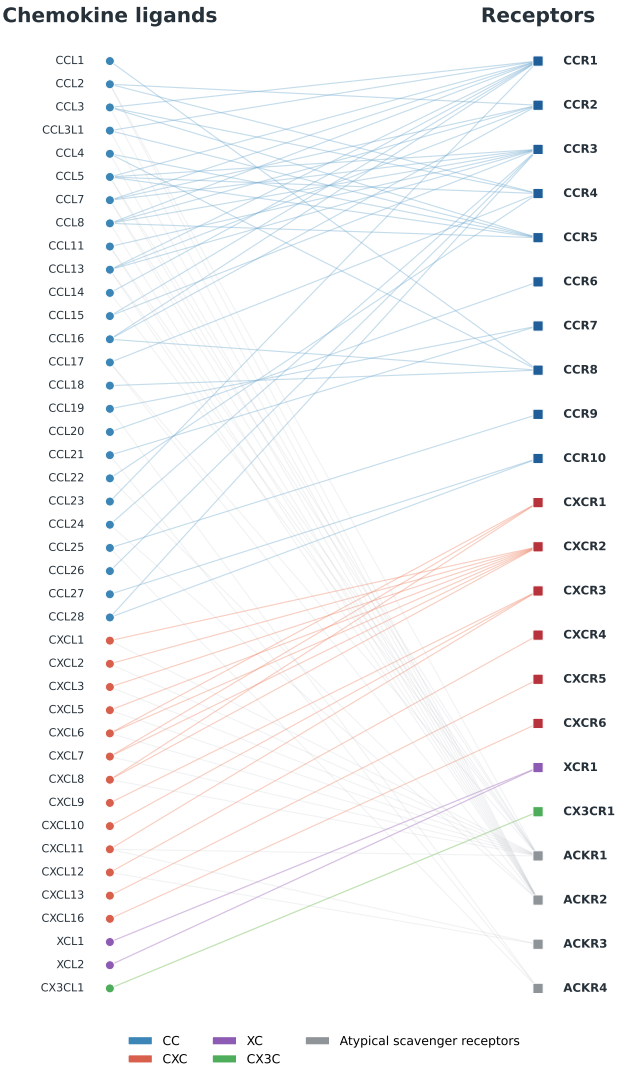

Figure 1.5: The human chemokine-receptor interaction network. Bipartite network representation showing chemokine ligands on the left, drawn as circles, connected to their cognate receptors on the right, drawn as squares. CC chemokines and receptors are shown in blue, CXC in red, XC in purple, and CX3C in green. Atypical scavenger chemokine receptors ACKR1–4 are shown in grey. Ligand and receptor nodes were defined from the IUPHAR/BPS Guide to PHARMACOLOGY chemokine-receptor interaction set, and edges represent activating interactions from the same source[61].

may mask functional specialisation, and the system may be less redundant than its binding topology suggests[88]. The degree of apparent promiscuity varies across the system. Some interactions are highly specific, such as CX3CL1 signalling exclusively through CX3CR1 and CXCL12 primarily through CXCR4, while others are highly

promiscuous[61,77]. The expression of multiple chemokine receptors by each leukocyte class further amplifies the connectivity of the system, as a single immune cell can respond to many different chemokines through potentially many different receptors[57,58].

1.4.4 Atypical chemokine receptors

In addition to the 18 conventional signalling receptors, the chemokine system includes four atypical chemokine receptors, ACKR1–4, that act as scavenger receptors and modulate chemokine availability without promoting chemotaxis[58,61]. These bind chemokines but fail to couple to G proteins due to alterations in the conserved DRYLAIV motif, Asp-Arg-Tyr-Leu-Ala-Ile-Val[61].

ACKR1, formerly DARC, is expressed on erythrocytes and endothelial cells, binds a broad range of CC and CXC chemokines, and buffers circulating chemokine concentrations[61]. ACKR1 may also mediate transcytosis by transporting chemokines through endothelial cells from the tissue-facing basolateral membrane to the blood-facing apical membrane[61]. ACKR2 (D6) is expressed on lymphatic endothelium and internalises inflammatory CC chemokines[61]. ACKR3 (CXCR7) binds CXCL11 and CXCL12 and can signal through β-arrestin pathways[61,77]. ACKR4 (CCRL1) scavenges homeostatic chemokines CCL19, CCL21, and CCL25[61]. The atypical receptors shape chemokine gradients and availability without directly promoting migration; ACKR2-deficient mice show enhanced inflammatory responses[61].

1.4.5 Chemokine gradients and glycosaminoglycan interactions

Chemokine function in vivo depends on the formation of concentration gradients that guide directional leukocyte migration[91,92]. Soluble chemokines alone may be

insufficient for directed migration, as chemokines can require immobilisation on cell surfaces and extracellular matrix to establish the spatial information required for chemotaxis[57,91,92]. This immobilisation is mediated by interactions between positively charged residues on the chemokine surface, typically in the 40s loop and C-terminal helix, and negatively charged glycosaminoglycans, which are sulfated carbohydrate chains on proteoglycans. These glycosaminoglycans (GAGs), including heparan sulfate, are present on endothelial and stromal cell surfaces[57,91,92].

There is evidence to suggest GAG binding is functionally essential for some chemokines. Chemokine mutants that retain receptor binding but lack GAG affinity fail to recruit leukocytes in vivo[91]. The GAG-binding surface overlaps with the dimerisation interface for many chemokines, linking oligomerisation to gradient formation: monomeric chemokines may bind receptors, but oligomeric forms are required for effective GAG presentation and in vivo function[85,91]. The composition and density of GAGs vary across tissues and are altered in inflammation, adding a spatial dimension to chemokine signalling that is not captured by expression data alone[92,93].

Chemokine gradients also undergo active remodelling by proteases, atypical receptors, and competing chemokines[61,92]. Proteolytic processing of chemokines by matrix metalloproteinases and dipeptidyl peptidase IV can convert agonists to antagonists or alter receptor selectivity[57,92]. These post-translational mechanisms mean that the functional chemokine environment in an inflamed tissue is more complex than the transcriptomic profile alone would suggest, a consideration relevant to any network analysis based on RNA expression data.

1.4.6 Evolution of the chemokine system

The chemokine system has evolved over approximately 500 million years, originating in early vertebrates[58,80]. Canonical chemokines and their receptors originate early

in the vertebrate lineage[80]. Chemokine receptors arise from a common ancestral GPCR that diversified through duplication events in the vertebrate stem group[80]. Whereas receptor diversity has remained relatively conserved, chemokine ligands have undergone successive expansions, first in jawed vertebrates, then in bony fish, and most dramatically in placental mammals[80]. Paralogs are genes that arise by duplication within a lineage and then diverge. In this context, CXCL1–3, CXCL5/6, CXCL16, CCL3, CCL18, and CCL25 refer to chemokine ligand genes that were generated or retained through mammalian lineage-specific duplication and innovation[80]. These paralogs increased the number of mammalian chemokine inputs without a comparable increase in conventional chemokine receptor number, creating conditions in which several ligands can converge on the same receptor.

This ligand-heavy organisation provides an evolutionary mechanism for the extensive ligand-receptor promiscuity and redundancy observed in human inflammatory disease[58,80]. Inflammatory chemokine genes form two major human clusters: CXC chemokines on chromosome 4 and CC chemokines on chromosome 17[58].

The selective pressures driving chemokine ligand expansion likely include host-pathogen coevolution[94,95]. Multiple mammalian viruses encode proteins that bind and sequester host chemokines, including poxviruses, herpesviruses, and cytomegaloviruses[94]. The myxoma virus protein M-T7 binds a broad range of CC, CXC, and C chemokines, while the variola virus protein CrmB binds both chemokines and TNF[94]. These viral proteins typically bind multiple chemokines, suggesting that effective immune evasion requires broad-spectrum neutralisation of the host chemokine response[94,95]. If viral immune evasion proteins selected for chemokine diversification as a host countermeasure, then the ligand expansions observed in the mammalian lineage may represent an adaptive response that makes it harder for any single viral protein to disable the entire chemokine system[80,94]. This arms race model predicts the redundant

network architecture observed in human disease and explains why parasites such as ticks have also evolved multi-chemokine inhibitors: both exploit the same vulnerability in their hosts[6,95].

The NF-κB and STAT axes are signalling categories rather than single evolutionary families, so their evolutionary histories cannot be compared with chemokines as a simple one-to-one gene-family contrast. The relevant comparison for this thesis is narrower. TNF and TNF receptor superfamilies expanded during vertebrate evolution, but they did not produce the mammalian chemokine-like pattern of a large ligand repertoire converging on a smaller set of conventional receptors[14,96]. IL-1 receptor genes and several IL-1 ligand genes also arose through ancient duplication events[97].

1.4.7 Chemokines in immune-mediated inflammatory disease

By directing leukocyte trafficking, chemokines contribute to the pathogenesis of atherosclerosis[18,21], type 1 diabetes[25], rheumatoid arthritis[45,55], inflammatory bowel disease[51,52], idiopathic pulmonary fibrosis[31], systemic lupus[39], psoriasis[33], and Sjögren's syndrome[42,54]. Inflamed tissues express multiple chemokines across classes. In rheumatoid arthritis, 15 CC and 11 CXC chemokines are expressed in joint tissues[45]. The human atherosclerotic plaque expresses 16 CC and 7 CXC chemokines[21,98]. This co-expression of many chemokines with overlapping receptor specificities creates a signalling environment in which blockade of any single chemokine or receptor is unlikely to prevent leukocyte recruitment, as alternative chemokine-receptor pathways remain available[86,87].

1.5 Therapeutic targeting of cytokine networks

1.5.1 NF-κB and STAT pathway inhibitors

The NF-κB and STAT signalling pathways have been the primary focus of therapeutic development[9,11]. Therapies such as TNF blockers, including infliximab, adalimumab, and etanercept, IL-6 receptor antagonists such as tocilizumab, and JAK inhibitors such as tofacitinib and baricitinib, are established treatments for rheumatoid arthritis, inflammatory bowel disease, and psoriasis[9–11]. The CANTOS trial demonstrated that canakinumab, a monoclonal antibody targeting IL-1β in the NF-κB pathway, reduces recurrent cardiovascular events in patients with atherosclerosis, providing proof-of-concept that targeting a single cytokine in this axis can produce clinical benefit even in a disease where chemokines are also heavily implicated[99]. IL-17A inhibitors such as secukinumab have shown high efficacy in psoriasis, with complete skin clearance in a substantial fraction of patients[35]. Anifrolumab, which blocks the type I interferon receptor in the STAT axis, is approved for systemic lupus erythematosus[37].

Despite these successes, drugs that block NF-κB and STAT cytokine networks have limited therapeutic efficacy across the broader patient population[10,19,20]. Approximately 30–40% of patients with rheumatoid arthritis discontinue anti-TNF treatment because of primary failure, secondary loss of response, or intolerance[19]. With intensive treatment, remission was achieved by 49% of patients with early rheumatoid arthritis and 19% with established rheumatoid arthritis[20]. Long-term immunosuppression carries risks of serious infection and malignancy[10,11]. The incomplete efficacy of existing therapies motivates the search for alternative therapeutic targets, including the chemokine system[7,86].

1.5.2 Failure of chemokine-targeted therapies

Despite their centrality in immune-mediated inflammatory disease, chemokine-receptor interactions have been challenging drug targets[7,8,45]. Numerous antagonists have entered clinical trials with limited success in inflammatory indications[7,8,45,100]. **Table 1.3** summarises selected chemokine receptor antagonists that have failed in clinical trials for inflammatory diseases. This pattern of failure is a major driver for seeking alternative approaches to tackle immune-mediated inflammatory diseases[8,45,86].

Table 1.3: Selected chemokine receptor antagonist programmes that did not yield approved therapies for the listed indications. Trial stages and reported outcomes were obtained from cited reviews and primary trial reports[7,8,45,100].

Compound	Target	Indication	Outcome
MLN1202	CCR2	Rheumatoid arthritis	No efficacy in Phase II
MK-0812	CCR2	Rheumatoid arthritis	No efficacy in Phase II
GSK766994	CCR3	Asthma and allergic rhinitis	Terminated in Phase II
Maraviroc	CCR5	Rheumatoid arthritis	No efficacy in Phase II
Vercirnon (CCX282-B; GSK1605786)	CCR9	Crohn's disease	Did not meet the primary endpoint in Phase III
SB-656933	CXCR2	Ulcerative colitis	Terminated; efficacy not assessed[101]
AZD5069	CXCR2	Asthma	No efficacy in Phase IIb[102]
Navarixin	CXCR2	COPD	No efficacy in Phase II
AMG487	CXCR3	Psoriasis	No efficacy in Phase II

However, the success of target-specific antagonists in non-inflammatory settings argues that clinical failures in immunoinflammation do not reflect target intractability[7,8]. **Table 1.4** lists FDA-approved drugs targeting chemokine receptors. Maraviroc blocks CCR5 to prevent HIV entry, plerixafor antagonises CXCR4 to mobilise haematopoietic stem cells, and mogamulizumab depletes CCR4-expressing malignant T-cells in cutaneous T-cell lymphoma[8,95]. In each case, the therapeutic mechanism depends on a single receptor-ligand interaction that is not compensated by alternative pathways:

HIV requires CCR5 for cell entry, stem cell retention depends on the CXCL12-CXCR4 axis, and the malignant T-cell clone expresses CCR4 as a marker[7,8]. Maraviroc failed when tested in rheumatoid arthritis despite its success in HIV, suggesting that the inflammatory context, where multiple chemokines and receptors are co-expressed, presents fundamentally different challenges from the single-pathway dependencies exploited in non-inflammatory settings[8,45].

Table 1.4: FDA-approved drugs targeting chemokine receptors. None are approved for immune-mediated inflammatory diseases.

Drug	Target	Modality	Indication	Year
Maraviroc	CCR5	Small molecule	HIV-1 infection	2007
Plerixafor	CXCR4	Small molecule	Stem cell mobilisation	2008
Mogamulizumab	CCR4	Antibody	Cutaneous T-cell lymphoma	2018
Motixafortide	CXCR4	Peptide	Stem cell mobilisation	2023
Mavorixafor	CXCR4	Small molecule	WHIM syndrome	2024

Several hypotheses have been advanced to explain the consistent failure of chemokine receptor antagonists in inflammatory disease[7,8]. Here I refer to three explanations for this pattern: the "wrong target" hypothesis, the "wrong drug" hypothesis, and the "wrong approach" hypothesis. The "wrong target" hypothesis proposes that individual clinical programmes selected targets that were not sufficiently relevant to the disease in question[7]. This explanation is difficult to sustain given the breadth of targets tested: CCR1, CCR2, CCR3, CCR5, CCR9, CXCR1, CXCR2, and CXCR3 have all been targeted across different diseases, and all have failed[7,8]. The "wrong drug" hypothesis suggests that individual compounds had insufficient receptor occupancy, selectivity, or pharmacokinetic properties[8]. While this may explain individual programme failures, it cannot explain the systematic pattern across chemically diverse compounds and modalities[7]. The CCR9 antagonist vercirnon, previously designated CCX282-B and GSK1605786, failed to meet the primary endpoint in a Phase III trial for Crohn's disease[100].

The "wrong approach" hypothesis proposes that the chemokine system has network-level properties that render single-target blockade inherently ineffective[7,86]. If multiple chemokines can activate the same receptor, or multiple receptors on the same cell can be activated by overlapping chemokine sets, then blocking one pathway leaves alternative routes for leukocyte recruitment intact[86,87]. This hypothesis predicts that the pattern of failure should be consistent across targets and diseases, which is observed, and that effective intervention should require multi-target approaches, which is supported by the efficacy of multi-chemokine inhibitors in animal models[95,103]. Testing this hypothesis formally requires quantitative analysis of chemokine network architecture in human disease tissues.

1.6 System complexity and network robustness

1.6.1 Complexity in the chemokine system

One explanation for therapeutic failure is system complexity, defined by "the quantity, the variety and the interconnectivity of its constituent elements"[104]. Complex systems have emergent properties that arise from interactions between components and cannot be predicted from the properties of individual components alone[3,105]. Several lines of evidence indicate that the chemokine system in immune-mediated inflammatory diseases is complex in this sense. In inflamed tissues, complexity arises from the expression of many chemokines across classes[21,33,45,51], the diversity of receptor expression on immune cells[58], and the high interconnectivity created by promiscuous ligand-receptor binding[58,61]. These interactions can cooperatively or synergistically modulate immune cell trafficking in ways that cannot be predicted from studying individual ligand-receptor pairs[84,85]. The chemokine system therefore meets the criteria for a complex system whose behaviour must be analysed at the network level rather

than the level of individual molecular interactions.

1.6.2 Biological robustness

A key emergent property of evolved complex systems is robustness, the capacity to maintain function despite perturbation[3,105,106]. Robustness arises through evolution to ensure survival under perturbations caused by mutation, environmental fluctuation, and stochastic noise[3,106]. Kitano identified four principal mechanisms by which biological systems achieve robustness[3]. Functional redundancy occurs where multiple components have overlapping functions, so that loss of one is compensated by another[3,4]. Modularity compartmentalises the system so that failure in one module does not propagate to others[3]. Feedback control allows systems to sense and correct perturbations[3,4]. Structural stability refers to the capacity of a system to maintain its qualitative behaviour despite parameter changes[3]. In the chemokine system, functional redundancy through overlapping ligand-receptor interactions is the most commonly invoked mechanism[57,87].

Multiple observations suggest that human chemokine signalling is robust[87,106]. The system has evolved over approximately 500 million years, suggesting sustained selective pressure for its maintenance[58,80]. Selective pharmacologic perturbations fail clinically, as documented across multiple targets and diseases[7,8,45,100]. Loss-of-function variants in chemokine receptors occur in healthy human populations without causing immunodeficiency: the CCR5-Δ32 allele occurs at frequencies of 4.4–14% across surveyed European populations[107], and the CX3CR1-M280 variant is associated with reduced atherosclerosis risk rather than immune impairment[108]. These observations suggest that the loss of individual chemokine receptors can be tolerated without compromising immune function, consistent with network-level compensation. Parasites have independently evolved multi-chemokine inhibitors, implying that multiple simultaneous

hits are required to disable the system[6,94,109]. Reported parasite-derived inhibitors outside the chemokine system include an IL-13 inhibitor and an IL-10 receptor antagonist[110,111]. The contrast between these examples and multi-target parasite strategies against chemokines suggests that the chemokine system may be unusually robust among cytokine systems.

1.6.3 The redundancy debate

The hypothesis that robustness in CK:CKR networks arises from redundancy, defined as overlapping ligand-receptor usage that drives immune cell migration, remains controversial[7,87]. The strongest arguments against redundancy come from mouse genetic studies. Dyer et al. showed that deletion of individual chemokine receptors in mice produced distinct phenotypes in models of peritoneal inflammation, with different receptors required for recruitment of different leukocyte subsets at different time points[88]. Metzemaekers et al. demonstrated that the three CXCR3 ligands, CXCL9, CXCL10, and CXCL11, have non-overlapping functions in T-cell recruitment despite binding the same receptor[112]. Ridley et al. showed that glycosaminoglycans can decode complex chemokine signals to produce distinct cellular responses, arguing that apparent redundancy masks context-dependent specificity[93].

These studies provide evidence for functional specialisation, but they were conducted in inbred mice maintained under controlled laboratory conditions[88,113]. Inbred mice lack the genetic and immune-cell diversity of outbred populations[114]. Pathogen-free housing limits immune experience, resulting in deficient effector and memory T-cell pools[115]. Studies using genetically diverse collaborative cross mice have revealed immune phenotypes absent from standard inbred strains, suggesting that genetic diversity affects immune network properties[114]. Humans, in contrast, live in pathogen-rich environments[116] and are genetically heterogeneous, with numerous genetic variants

in chemokine and receptor genes[117]. These differences may explain why redundancy is less evident in mouse models but may be present in human disease. Chemokine compensation favoured in genetically diverse hosts with repeated microbial exposure may be underestimated in experimental systems with fixed genetic backgrounds, restricted immune experience, and reduced cellular heterogeneity[86,87]. Resolving this debate requires direct analysis of chemokine network properties in human disease tissues.

1.7 Network biology and graph-theoretic analysis

Testing whether the chemokine system is redundant and robust requires a quantitative framework. Network biology provides such a framework by representing molecular systems as graphs in which nodes are components and edges are interactions[118]. The arrangement of these interactions determines redundancy, modularity, and robustness at the system level, beyond properties measurable from isolated components[3,118]. This section introduces the graph-theoretic concepts used in this thesis.

1.7.1 Bipartite networks

Ligand-receptor signalling systems are naturally represented as bipartite networks, in which nodes are divided into two disjoint sets, with edges connecting only nodes from different sets[119]. In a cytokine signalling network, one set contains ligands and the other contains receptors, with edges representing activating interactions. Bipartite networks have been studied extensively in ecology, where they describe mutualisms between plants and their pollinators or seed dispersers[119,120]. The mathematical properties of bipartite graphs, including metrics for redundancy and robustness, are well established and directly applicable to ligand-receptor systems[119].

The degree of a node in a bipartite network is the number of edges connecting it to nodes in the other set. In a ligand-receptor network, ligand degree might measure the number of receptors a ligand activates, and receptor degree might measure the number of ligands that activate a receptor, provided edges are defined as activating ligand-receptor interactions[118]. Degree distribution, the pattern of how connections are distributed across nodes, determines many global network properties including vulnerability to perturbation[121,122].

1.7.2 Measuring network redundancy

The minimum edge cover is the smallest set of edges in which every node is incident to at least one edge[123]. This thesis defines edges beyond that minimum as redundant because they are not required for node coverage.

The relationship between the minimum edge cover and the maximum matching in a bipartite graph is given by the Gallai identity: $|V| = |E_{\min}| + M$, where $|V|$ is the number of nodes, $|E_{\min}|$ is the minimum edge cover size, and M is the maximum matching size[123]. The maximum matching is the largest set of edges in which no two edges share a node, representing the maximum number of one-to-one ligand-receptor pairings. If E is the complete set of observed edges, the normalised edge redundancy index is $R = 1 - |E_{\min}|/|E|$, where $|E|$ is the observed edge count and $|E_{\min}|$ is the minimum edge-cover count. The index ranges from 0, where every edge is necessary, to values approaching 1, where many edges exceed the minimum required for node coverage, and therefore allows comparison across networks of different sizes.

1.7.3 Measuring network robustness

Network robustness quantifies the capacity of a network to maintain function under progressive perturbation[121,122]. The foundational observation in network science is that

scale-free networks, in which most nodes have few connections but a small number of hubs have many, are robust to random node failure but vulnerable to targeted attack on hubs[121]. This finding established that network topology determines vulnerability to disruption[121,122].

Two complementary approaches to measuring robustness are used in this thesis. The first, secondary extinction analysis, was developed in ecology to assess the vulnerability of mutualistic networks to species loss[120,124]. In this approach, nodes from one set of the bipartite network are sequentially removed. A node in the other set is considered lost when all its connections to the first set are severed. The area under the curve plotting the fraction of surviving secondary nodes against the fraction of primary nodes removed quantifies functional robustness[120]. Applied to ligand-receptor networks, this approach models the consequence of progressive ligand loss on receptor activation: a receptor remains functional as long as at least one activating ligand persists, but becomes silent when all its ligands are removed.

The second approach, structural robustness analysis, measures how rapidly a network fragments under node removal[125]. At each step of a removal sequence, the fraction of nodes remaining in the largest connected component is recorded. The area under this curve, normalised to the initial node count, quantifies structural robustness[125]. Networks that maintain large connected components under progressive attack are structurally robust; those that fragment rapidly are vulnerable[122,125].

1.7.4 Prior network-level analyses of chemokine signalling

Several studies have characterised the chemokine system at the network level. Von Hundelshausen et al. mapped the chemokine interactome by systematically measuring chemokine-chemokine interactions using immunoligand blotting and surface plasmon resonance, revealing that heteromeric chemokine interactions can enhance

or inhibit receptor activation depending on the chemokine class[84]. Kleist et al. constructed a comprehensive interaction network of over 100 experimentally validated chemokine-GPCR interactions and used sequence and structural analysis to identify the molecular determinants of selectivity and promiscuity, showing that a minimal set of conserved residues encodes general chemokine-receptor recognition while variable residues customise pair-specific binding[83]. Bhattacharya and Kawamura framed the chemokine system explicitly as a network and argued that its redundancy necessitates multi-target inhibition for therapeutic efficacy[86].

These studies established that the chemokine system has network-level organisation with extensive promiscuity, but none quantified the topological properties of chemokine networks relative to other cytokine networks. The question of whether chemokine networks are more redundant and robust than NF-κB or STAT networks, and whether this might explain the differential success of therapeutics targeting these axes, has not been addressed.

1.7.5 Application to cytokine signalling

Applying the graph-theoretic measures described above to cytokine signalling networks tests the redundancy hypothesis directly. If chemokine-receptor networks are more redundant than NF-κB or STAT networks, they should have higher edge redundancy indices, reflecting greater excess of edges beyond the minimum required for connectivity. If this redundancy translates into robustness, chemokine networks should maintain receptor activation and structural integrity longer under simulated perturbation than NF-κB or STAT networks. Comparing human expression-derived networks across diseases and signalling axes can establish whether chemokine topology differs from that of cytokine pathways with successful therapies.

1.8 Tick evasins as multi-chemokine inhibitors

If chemokine network redundancy underlies therapeutic failure, effective inhibition requires simultaneous neutralisation of multiple chemokines. Nature has solved this problem independently in at least two lineages: blood-feeding ticks and viruses have both evolved proteins that bind and neutralise multiple chemokines[6,94,95]. The convergent evolution of multi-chemokine inhibitors from unrelated organisms suggests that single-target blockade is fundamentally insufficient to disable the chemokine system[86,95].

Ticks are obligate blood-feeding arachnids that remain attached to a host for days or weeks[6,126]. Neutrophils, macrophages, and lymphocytes accumulate at the feeding site and can disrupt attachment or blood acquisition[6]. Salivary immunomodulators counter this response; they include more than 50 reported evasins from several tick genera[6,126–128].

Evasins comprise two structural classes distinguished by chemokine preference[6,126]. Class A proteins contain four or five disulfide bonds and recognise CC chemokines, with Evasin-1 and Evasin-4 from *Rhipicephalus sanguineus* as established examples. Early biochemical and structural studies characterised Evasin-1 and the Class B protein Evasin-3[129]. The Evasin-1–CCL3 crystal structure places the evasin against the chemokine N-loop and 40s loop, which overlap the receptor-binding surface[130]. Class B proteins instead recognise CXC chemokines through a three-disulfide knottin fold with high thermal and proteolytic stability[6,131].

Binding breadth also varies within a structural class: six CC chemokine partners have been reported for Evasin-1 and approximately 20 for Evasin-4[6]. Two hypotheses could explain how one evasin accommodates several partners. Conformational plasticity

proposes that one binding surface changes shape to complement different chemokines, whereas differential positioning assigns different partners to separate surface patches. These alternatives imply different therapeutic designs. Plastic interfaces could support compact multi-target binders, while multiple recognition patches could favour multi-site constructs. Evasin-derived peptides and broadly neutralising antibodies show that broad chemokine recognition can be engineered[132–134].

Experimental structures are too sparse to compare the same evasin across several chemokine partners because these complexes are difficult to crystallise or analyse by NMR[6]. Available data include Evasin-1 bound to CCL3 and the unbound X-ray structure of EVA3, but no EVA3–CXCL8 complex[130,131]. Computational prediction therefore provides a way to test whether binding interfaces vary between partners.

Mouse studies provide evidence that broad chemokine binding can suppress inflammation[95,103,129]. Evasin-1 reduced inflammation in peritonitis and lung-injury models and lowered mortality after intestinal ischaemia-reperfusion[129]. Evasin-4 limited myocardial injury after infarction and reduced disease severity in contact-hypersensitivity and colitis models[103]. The contrast with trials of selective antagonists provides a rationale for testing whether broader blockade can overcome network redundancy[86,95].

1.8.1 Multi-target therapeutic approaches

The efficacy of natural multi-chemokine inhibitors motivates the development of therapeutic agents that replicate this strategy[86,95]. Several approaches have been pursued.

Broadly neutralising antibodies have been engineered using yeast surface display to bind multiple chemokines, exploiting the structural conservation of chemokines to identify antibody variants that cross-react with diverse family members[134]. This

approach demonstrated efficacy in a mouse model of arthritis, supporting the principle that multi-chemokine targeting can suppress inflammation in vivo[134].

Evasin-derived peptides offer another approach. Vales et al. used phage display screening of peptide libraries derived from class A evasins, performed in parallel against multiple chemokines, to identify short peptides with broad-spectrum chemokine binding[132]. The lead peptide HD2, derived from the class A evasin EVA4, binds chemokines from both CC and CXC classes despite class A evasins normally binding only CC chemokines[132]. AlphaFold 2 modelling suggested that the peptide occludes distinct receptor-binding regions in CC and CXC chemokines, with two conserved motifs contributing ionic and hydrophobic interactions respectively[132]. Building on this work, Kryukova et al. developed a combinatorial saturation mutagenesis optimisation strategy to enhance the potency and breadth of the HD2 peptide[133]. By combining improving point mutations identified through saturation mutagenesis phage display, they identified combinatorially mutated peptides with significantly enhanced inhibitory activity against multiple inflammatory chemokine pools[133]. These peptides can be engineered into peptibody fusions with IgG1-Fc domains for improved pharmacokinetics and half-life[132]. Despite their non-immunoglobulin origin, evasins may still be immunogenic when used therapeutically[133]. Short peptides derived from evasins are less likely to elicit immune responses than full-length evasin proteins and can be produced synthetically at lower cost[132,133].

Chemokine-derived peptides that interfere with chemokine oligomerisation are a third strategy: peptides based on the CCL5 N-terminus can disrupt chemokine heteromerisation and synergy[84]. As mentioned previously, no multi-target chemokine inhibitor has reached clinical use for inflammatory disease[7,86]. Understanding the structural basis by which evasins achieve broad-spectrum binding could inform the rational design of such therapeutics.

1.9 Single-cell RNA sequencing in immunology

Single-cell RNA sequencing (scRNA-seq) captures the transcriptome of individual cells, revealing cellular heterogeneity that bulk RNA sequencing obscures[135]. Bulk RNA-seq averages gene expression across thousands to millions of cells in a tissue sample, producing a single expression profile that conflates the contributions of different cell types[135]. In inflamed tissues, where diverse immune and stromal cell populations coexist, this averaging masks the cellular sources of cytokine production and receptor expression. For example, a cytokine that appears moderately expressed in bulk may in fact be highly expressed by a rare immune cell subpopulation and absent from the majority of cells. Single-cell analysis is needed to distinguish between these scenarios.

scRNA-seq resolves this limitation by generating expression profiles for individual cells[135]. The application of scRNA-seq to inflamed tissues has revealed that cells appearing homogeneous by surface marker expression can display transcriptional diversity, with subpopulations exhibiting distinct functional states[135]. This resolution enables the identification of which cell types produce specific cytokine ligands and which immune cells express their cognate receptors, allowing construction of cell-type-resolved interaction networks that reflect the molecular organisation of signalling as it operates in diseased tissue[135,136].

Cell-cell communication inference from scRNA-seq data has become a major analytical application[136,137]. Tools such as CellChat construct ligand-receptor interaction networks from single-cell expression profiles, incorporating curated databases of known interactions to predict signalling between cell types[136]. Systematic comparisons of cell-cell communication inference methods have found that both the choice of interaction database and the analytical method influence the predicted interactions, and that methods show varying sensitivity to tissue-specific and pathway-specific sig-

nals[137]. These considerations motivate the use of experimentally validated interaction databases, such as the International Union of Pharmacology (IUPHAR) database, rather than computationally inferred interactions[61].

Several computational methods support downstream analysis of scRNA-seq data[135]. Dimensionality reduction techniques such as UMAP enable visualisation of high-dimensional transcriptomic data in two dimensions[138]. Cell type annotation can be performed using reference-based classifiers such as CellTypist, which uses machine learning trained on a comprehensive immune cell atlas to assign cells to known categories across tissues[139]. Cross-tissue applicability is relevant for this thesis because the twelve diseases examined span different organ systems, and consistent cell type annotation across tissues is necessary for meaningful comparison.

scRNA-seq has specific technical limitations that affect the analysis of cytokine networks. Most droplet-based platforms isolate single cells in microfluidic droplets together with barcoded beads. During library preparation, transcripts from each cell receive a cell barcode and unique molecular identifiers (UMIs) before pooled sequencing, allowing reads to be assigned back to their cell of origin. These platforms dominate current studies, but they have limited sensitivity for lowly expressed transcripts, and the sparsity of single-cell data means that a gene may go undetected in a cell even when it is expressed[135,140].

Spatial single-cell RNA sequencing adds cell-position information to single-cell expression measurement. This is relevant for chemokine biology because leukocyte recruitment depends on concentration gradients and tissue proximity, not only on whether a ligand and receptor are detectable somewhere in a dissociated sample[57,91,92]. Spatial scRNA-seq therefore allows ligand-receptor analysis to ask whether ligand-expressing cells and receptor-expressing immune cells are positioned near each other

in diseased tissue.

A further consideration is that transcript-based network construction uses RNA as a proxy for protein expression and function. Transcript and protein abundance correlate moderately[141]. Post-translational mechanisms including proteolytic processing, glycosylation, and protein stability introduce variation that is not captured by transcriptomic data[57,141]. Matrix metalloproteinases and dipeptidyl peptidase IV can modify chemokine activity after translation by converting receptor agonists into antagonists or changing receptor selectivity[57]. These modifications are not represented in transcript-based networks. Despite these limitations, scRNA-seq provides the most direct available approach for characterising cytokine signalling networks as they operate in human disease tissues, and it avoids the assumptions inherent in alternative methods such as computational deconvolution of bulk data[135].

1.10 Computational structural biology and AlphaFold

Understanding how evasins achieve broad-spectrum chemokine binding requires structural characterisation of multiple evasin-chemokine complexes. Experimental structure determination by X-ray crystallography or NMR is slow and technically demanding for each individual complex. Fully mapping the evasin:chemokine, chemokine:chemokine, and chemokine:receptor interaction sets would require hundreds of structures, making experimental resolution of the full network infeasible. Computational structure prediction offers an alternative.

AlphaFold is a deep learning system developed by DeepMind that predicts protein structure from amino acid sequence[142]. AlphaFold 2 achieved a median domain Global Distance Test score of 92.4 in CASP14[143]. The model uses a multiple sequence alignment and structural templates as inputs, processing them through an attention-

based neural network that outputs three-dimensional coordinates and per-residue confidence scores[142]. The predicted local distance difference test (pLDDT) estimates local confidence on a scale of 0–100[142].

AlphaFold-Multimer extended the AlphaFold 2 architecture to predict the structures of protein complexes[144]. On a benchmark of 17 heterodimers, AlphaFold-Multimer produced models of at least medium accuracy for 13 targets, including high-accuracy models for seven[144]. The CASP15-CAPRI experiment confirmed this progress: high-quality complex models were produced for approximately 40% of targets, compared with 8% two years earlier[145]. Performance remained limited for antibody-antigen complexes, where evolutionary relationships between binding partners are absent, and for complexes involving substantial conformational change[145].

AlphaFold 3, released in 2024, uses a diffusion-based architecture that jointly predicts the structures of protein complexes together with nucleic acids, small molecules, and ions[146]. For protein-protein interactions, AlphaFold 3 achieves accuracy comparable to or exceeding AlphaFold-Multimer, with particular improvements for complexes involving post-translational modifications and ligand binding[146]. The predicted aligned error (PAE) provides a measure of relative domain positioning, complementing the per-residue pLDDT scores for assessing prediction quality[146].

The application of AlphaFold to chemokine biology is methodologically suited to the questions addressed in this thesis. Chemokine-evasin interactions involve well-folded, disulfide-stabilised proteins with substantial sequence homology within their respective families, features that favour accurate prediction[6,146]. Multiple sequence alignments for both chemokines and evasins are readily constructed from vertebrate and tick genomes, providing the evolutionary signal on which AlphaFold depends[142]. By modelling the same evasin in complex with different chemokine partners, structural variation at the

binding interface can be systematically assessed across hundreds of interactions, a scale that is infeasible with experimental methods alone.

Alternative computational approaches to protein-protein interaction modelling include molecular docking and molecular dynamics simulation. Docking methods sample orientations and conformations of two proteins to find energetically favourable binding poses, but they typically require known structures of the individual partners and have difficulty modelling backbone flexibility at binding interfaces[147]. Molecular dynamics simulations can capture conformational changes but are computationally expensive and impractical for systematic analysis of hundreds of complexes[148]. AlphaFold 3 predicts both the individual protein structures and their complex from sequence alone, making it suited to large-scale analyses where experimental structures are unavailable for most interaction pairs[146].

AlphaFold predictions are not experimental observations and require validation against available experimental data[146]. Where crystal or NMR structures exist, predicted models can be assessed by root mean square deviation of atomic coordinates. Where experimental structures are unavailable, confidence metrics and consistency across related predictions provide indirect quality measures[142,146]. The validation strategy used in this thesis is described in Chapter 2 and applied in Chapter 5.

1.11 Thesis aims and objectives

This thesis tests two hypotheses. The first is that chemokine-receptor networks in human IMIDs are characterised by redundancy and robustness that exceeds that of NF-κB and STAT networks, providing a mechanistic explanation for the clinical failure of single-target chemokine inhibitors. If this hypothesis is supported by the analysis, chemokine networks should have higher edge redundancy indices than NF-κB and

STAT networks, reflecting greater excess of edges beyond the minimum required for connectivity. They should also maintain receptor activation and structural integrity longer under simulated perturbation. These differences should be consistent across diseases rather than specific to individual conditions.

The second hypothesis is that the broad specificity of tick evasins arises from conformational plasticity at the binding interface, enabling them to sterically occlude functional chemokine sites and overcome network robustness. If this hypothesis is supported by the analysis, the same evasin residues should be used to bind different chemokine partners, but those residues should adopt different conformations across complexes. Evasin-binding sites on chemokines should overlap with receptor-binding and dimerisation sites, consistent with a mechanism of steric occlusion.

To test these hypotheses, this thesis has three objectives. The first objective is to analyse single-cell RNA-sequencing data from twelve immune-mediated inflammatory diseases to characterise cytokine ligand and receptor expression across tissue cell types and immune cell populations (Chapter 3). The second objective is to construct and analyse cytokine-receptor interaction networks for the chemokine, NF-κB, and STAT signalling axes across diseases, quantifying their redundancy and robustness using network biology approaches (Chapter 4). Within the first and second objectives, public spatial scRNA-seq data from psoriasis skin and ulcerative colitis rectal tissue are used to ask whether the CK:CKR network pattern is also observed when cell-position information is used to restrict ligand-receptor interactions by tissue proximity[149,150]. The third objective is to investigate the structural basis of evasin-chemokine interactions using AlphaFold 3 to model 523 chemokine interactions and analyse the mechanisms by which evasins achieve broad-spectrum chemokine inhibition (Chapter 5).

Several aspects of the approach should be noted. The ligand-receptor networks

constructed in this thesis are unweighted: edges represent the presence of an activating interaction between an expressed ligand and an expressed receptor, but do not encode binding affinity, signalling potency, or expression magnitude. This design choice reflects the topological nature of the redundancy and robustness measures used, and the fact that potency values derived from receptor-transfected cell lines may not reflect in vivo behaviour. The scRNA-seq data are cross-sectional, representing a snapshot of cytokine expression at the time of tissue sampling rather than the temporal dynamics of inflammation. The primary twelve-disease scRNA-seq analysis used dissociated datasets, while chemokine function depends on localised concentration gradients shaped by glycosaminoglycan binding and tissue microanatomy[91]. Public spatial scRNA-seq datasets from psoriasis lesional skin and ulcerative colitis inflamed rectal tissue were therefore analysed because they reported gene expression and tissue position for individual cells[149,150]. The AlphaFold 3 structural predictions are computational models, not experimental structures, and require validation against available crystallographic and NMR data. These considerations are addressed in the methods and discussion chapters.

1.12 Thesis organisation

Chapter 1, this chapter, introduces the biological background, therapeutic context, and aims. It establishes the rationale for comparing chemokine signalling with the NF-κB and STAT axes.

Chapter 2 describes the computational and analytical methods used throughout the thesis, including the scRNA-seq processing pipeline, network construction and analysis methods, and the AlphaFold 3 structural prediction workflow. These methods support the expression, network, and structural analyses in the subsequent chapters.

Chapter 3 presents the analysis of scRNA-seq data from twelve immune-mediated inflammatory diseases. This chapter establishes the molecular foundation for network construction by characterising which cytokine ligands and receptors are expressed in diseased tissues and by which cell types. It also uses public spatial scRNA-seq datasets from psoriasis skin and ulcerative colitis rectal tissue to show which cytokine-axis genes could be analysed when tissue positions were available. The key observation from this chapter is the asymmetry between chemokine ligand and receptor diversity, which sets up the network analysis in the following chapter.

Chapter 4 constructs and analyses 36 cytokine-receptor interaction networks, comprising 12 diseases and 3 signalling axes. This chapter contains the central quantitative result of the thesis: the demonstration that chemokine networks are significantly more redundant and robust than NF-κB or STAT networks. The chapter also analyses spatial scRNA-seq networks using cell-position information from psoriasis lesional skin and ulcerative colitis inflamed rectal tissue.

Chapter 5 presents structural analysis of 523 chemokine interactions using AlphaFold 3, investigating how tick evasins achieve broad-spectrum chemokine binding. This chapter addresses the mechanism by which natural inhibitors overcome the network redundancy established in the preceding chapter.

Chapter 6 integrates the findings, discusses their implications for therapeutic development and the redundancy debate, and considers the evolutionary perspective and limitations of the work. It closes by identifying experimental and clinical tests of the network framework.

Chapter 2

Materials and Methods

2.1 Public databases

The analyses in this thesis drew on several public databases and data portals. This section summarises each resource and its role. Full URLs are given for reproducibility.

2.1.1 Gene Expression Omnibus

The Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (`https://www.ncbi.nlm.nih.gov/geo/`) is a public repository for high-throughput functional genomics data. Single-cell RNA-sequencing datasets used in this thesis were obtained from GEO series accessions as described in Section 2.2.

2.1.2 CELLxGENE Discover

CELLxGENE Discover (Chan Zuckerberg Initiative; `https://cellxgene.cziscience.com/`) is a portal for standardised single-cell transcriptomics data. The type 1 diabetes pancreatic islet dataset (GSE148073) was downloaded from CELLxGENE rather than from GEO.

2.1.3 GENCODE

GENCODE (`https://www.gencodegenes.org/`) provides comprehensive human gene annotation. The GENCODE release 46 GTF file (`https://www.gencodegenes.org/human/release_46.html`) was used to map Ensembl gene identifiers to gene symbols during data wrangling.

2.1.4 Ensembl

Ensembl (`https://www.ensembl.org/`) is a genome browser and annotation system that assigns stable Ensembl gene identifiers. Source scRNA-seq matrices reported genes using Ensembl identifiers where applicable; these identifiers were reconciled with gene symbols using the GENCODE annotation described above.

2.1.5 UniProt

The UniProt Knowledgebase (`https://www.uniprot.org/`) is the central resource for protein sequence and functional annotation. Human cytokine and receptor genes were identified by querying UniProt with keyword and Gene Ontology filters as described in Section 2.5. Protein sequences for structural modelling were obtained from UniProt. UniProt was also used to identify PDB entries associated with human chemokines for comparison with AlphaFold 3 models. Chemokine-chemokine interaction records were retrieved through PSICQUIC with UniProt as one of the interaction providers (Section 2.1.12).

2.1.6 Gene Ontology

The Gene Ontology resource (`https://geneontology.org/`) provides controlled vocabulary terms for gene function. Gene Ontology identifiers listed in Section 2.5 were

used as query filters in the UniProt search for cytokine and receptor genes.

2.1.7 Guide to Pharmacology (IUPHAR/BPS)

The IUPHAR/BPS Guide to Pharmacology (`https://www.guidetopharmacology.org/`) is a curated database of drug targets and ligand-target interactions. Ligand-target interaction tables were downloaded from the project download page (`https://www.guidetopharmacology.org/download.jsp`) and used to define network edges for cytokine-receptor network construction and for compiling activating chemokine:receptor pairs in the structural interaction set (Sections 2.7 and 2.10).

2.1.8 BioGRID

The Biological General Repository for Interaction Datasets (BioGRID; `https://thebiogrid.org/`) archives protein and genetic interaction records. Chemokine-chemokine physical interactions were retrieved through the PSICQUIC query interface with BioGRID as a provider (Section 2.1.12).

2.1.9 IntAct

IntAct (`https://www.ebi.ac.uk/intact/`) is a molecular interaction database maintained at EMBL-EBI. It was included as a PSICQUIC provider for chemokine-chemokine interaction retrieval (Section 2.1.12).

2.1.10 IMEx

The International Molecular Exchange (IMEx; `https://www.imexconsortium.org/`) consortium coordinates curated molecular interaction data. IMEx was included as a PSICQUIC provider for chemokine-chemokine interaction retrieval (Section 2.1.12).

2.1.11 Molecular INTeraction database (MINT)

MINT (`https://mint.bio.uniroma2.it/`) is a curated database of molecular interactions. MINT was included as a PSICQUIC provider for chemokine-chemokine interaction retrieval (Section 2.1.12).

2.1.12 PSICQUIC interaction providers

PSICQUIC (`https://psicquic.github.io/`) is a standardised web service protocol for querying multiple molecular interaction databases. PSICQUIC queries were made to the providers BioGRID, IntAct, IMEx, MINT, and UniProt; only interactions supported by NMR, X-ray crystallography, biolayer interferometry, or surface plasmon resonance were retained (Section 2.10).

2.1.13 Protein Data Bank

The Worldwide Protein Data Bank (wwPDB; `https://www.wwpdb.org/`) is the archive of three-dimensional structural data for biological macromolecules. Experimental structures deposited in the PDB were used to validate AlphaFold 3 models by RMSD and to select highest-resolution structures where multiple entries existed (Section 2.11).

2.1.14 AlphaFold 3 structure prediction server

Protein complex structures were generated using the AlphaFold 3 webserver (`https://alphafoldserver.com/`; DeepMind and Isomorphic Labs) as described in Section 2.11. The server is a public access point for AlphaFold 3 inference.

2.2 Single-cell transcriptomic data sources

Publicly available single-cell transcriptomic datasets were analysed. The main non-spatial analysis used single-cell RNA sequencing datasets from twelve immune-mediated inflammatory diseases. These datasets were selected because they span distinct organ systems, contribute substantially to global disease burden, and satisfy the inclusion criteria described in **Table 2.1**.

Table 2.1: Inclusion criteria for scRNA-seq dataset selection.

#	Criterion	Reason
1	Inflammatory disease	Ensures relevance to immune-mediated inflammatory disease and cytokine signalling
2	Primary tissue of interest	Enables analysis of cytokine networks at the site of inflammation rather than circulating cells
3	Single-cell RNA-seq	Provides cell-level resolution required for ligand-receptor network construction
4	No pre-enrichment for cell types	Avoids bias from physical sorting and ensures representation of all tissue-resident populations
5	Raw transcript counts available	Enables standardised quality control and normalisation across datasets

Carotid atherosclerosis data were obtained from GSE159677[151]. Type 1 diabetes pancreatic islet data were obtained from GSE148073 via CELLxGENE[25]. Crohn's disease and ulcerative colitis colonic mucosal biopsy data were obtained from GSE214695[152]. Idiopathic pulmonary fibrosis lung biopsy data were obtained from GSE135893[153]. Psoriasis skin biopsy data were obtained from GSE173706[154]. Systemic lupus erythematosus and discoid lupus erythematosus skin biopsy data were obtained from GSE179633[39]. Sjögren's syndrome salivary gland biopsy data were obtained from GSE272409[42]. Rheumatoid arthritis and psoriatic arthritis synovial biopsy data were obtained from GSE200815[46]. Endometriosis ectopic endometrium data were obtained from GSE213216[50].

The scRNA-seq datasets were downloaded from the Gene Expression Omnibus, with the exception of GSE148073 which was downloaded from CELLxGENE. GEO datasets were selected such that cells had not been physically pre-sorted, so that cytokine expression by all cell types could be assessed.

The datasets were not integrated into a single cell-level atlas before expression quantification. This was a deliberate choice because the tissues, sequencing protocols, disease states, and available controls differed between studies. Instead, each dataset was processed with the same pipeline and summarised at the disease level. This approach preserves the biological identity of each disease-tissue pair while allowing the same downstream rules to be applied to all conditions. It also avoids forcing cells from unrelated tissues into a shared low-dimensional space where batch correction could remove real disease- or tissue-specific signal. The unit of comparison in the network analysis was therefore the disease-level ligand-receptor repertoire rather than individual cells or clusters pooled across all studies.

Matched healthy controls were used only for analyses that required disease-versus-control comparison, such as fold change in ligand, receptor, target-gene, or cell-type abundance. They were not required for disease-network construction, because the network analysis asked which cytokine ligands and immune-cell receptors were expressed in diseased tissue itself. For diseases lacking a matched healthy control within the same study, disease-state expression could still be used to construct ligand-receptor networks, but fold-change interpretation was not performed.

Spatial network analysis extended the cytokine-receptor network analysis by requiring ligand-receptor edges to be observed within local tissue tiles. A public inventory of spatial scRNA-seq datasets and assay panels was manually screened to identify datasets with gene expression and tissue position for individual cells and sufficient measurement

of chemokine, NF-κB, and STAT cytokine-receptor genes. Two spatial scRNA-seq disease datasets satisfied these requirements: GSE314158, which sampled psoriasis skin, and GSE312415, which sampled rectal tissue from patients with ulcerative colitis[149,150]. Both datasets were generated with the CosMx 6K platform, a targeted spatial molecular imaging assay in which approximately 6 000 RNA targets are measured in intact tissue sections and assigned to segmented cells with field-of-view and coordinate metadata[149,150]. Because these datasets were generated with a targeted assay that measures a fixed gene panel rather than the whole transcriptome, panel coverage was assessed before network construction. Tiled local spatial network analysis was performed on lesional skin from patients with psoriasis and inflamed rectal tissue from patients with ulcerative colitis.

2.3 Data processing pipeline

The complete computational workflow is summarised in **Figure 2.1**. It links source-data preparation to cell classification, expression filtering, network construction, and downstream robustness analyses.

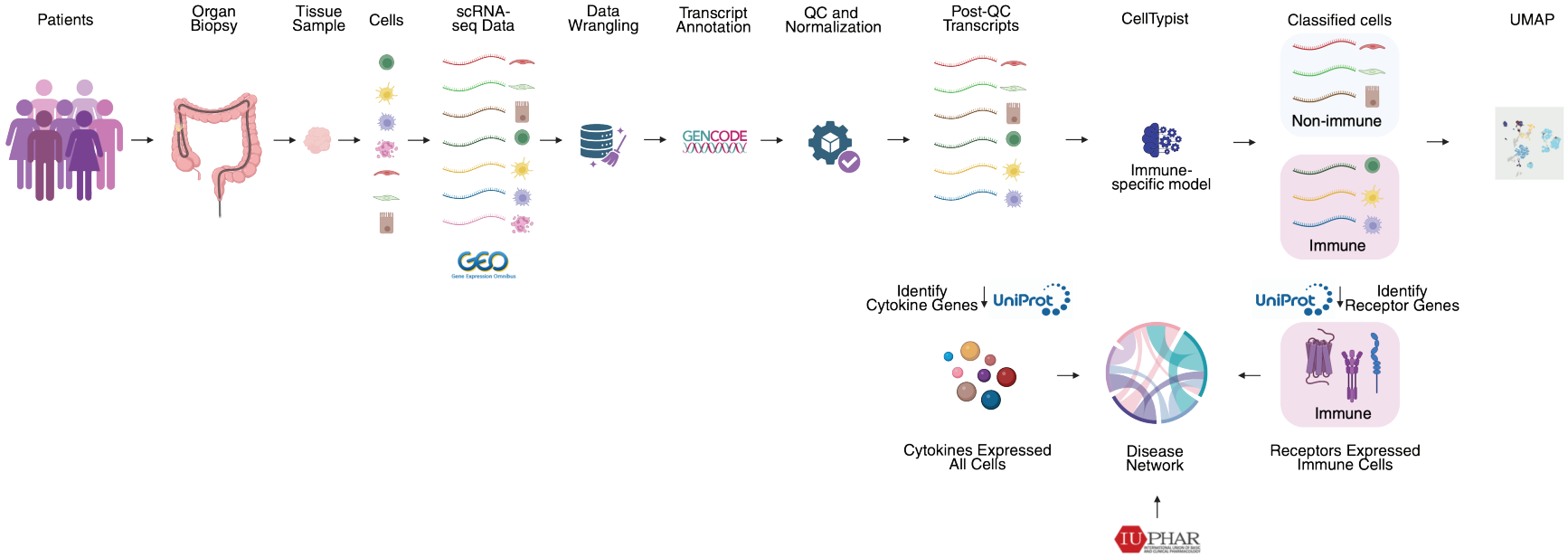

Figure 2.1: Overview of the computational pipeline for processing single-cell RNA sequencing data and constructing cytokine-receptor interaction networks. Patient organ biopsy, tissue sampling, cell isolation, and scRNA-seq data generation are shown for context; the work in this thesis began with acquired scRNA-seq datasets. These data were processed through data wrangling, transcript annotation, quality control, normalisation, immune and non-immune cell classification, UMAP visualisation, cytokine and receptor gene identification, disease-level expression quantification, integration of IUPHAR ligand-receptor interactions, disease network construction, and network analysis using redundancy and robustness measures.

2.3.1 Data wrangling and annotation

Raw RNA transcript abundances were extracted from the source datasets using Python scripts to parse sample matrix files, identify correct sample identifiers based on series matrix files, and import the output of different scRNA-seq platforms. Data from individual sample identifiers were combined using the Python package anndata to generate h5ad annotated data matrix files that stored cell barcodes in rows and genes in columns. To maintain consistency across datasets, Ensembl gene identifiers were converted to gene symbols using gencode.v46.annotation.gtf from GENCODE (`https://www.gencodegenes.org/human/release_46.html`).

For the spatial scRNA-seq datasets, expression matrices were joined to cell metadata, sample annotations, field-of-view identifiers, and field-of-view position files. Cell centre coordinates supplied with the CosMx 6K metadata were retained as global pixel coordinates. The micrometre-per-pixel scale was derived from the pixel and millimetre coordinates recorded for each field of view, and this scale was applied to the global cell-centre coordinates to generate tissue coordinates for tiling. Cells without usable coordinate or total-count metadata were excluded before local network construction.

2.3.2 Quality control

Cells with more than 10% mitochondrial transcripts were excluded, and cells with more than 10 000 total transcripts or genes were excluded as potential multiplets. Cells with fewer than 200 total transcripts or genes were excluded as low-complexity profiles. These thresholds were applied consistently across datasets.

2.3.3 Normalisation

Transcript counts were normalised to 10 000 total transcripts per cell, followed by log1p transformation such that each value x became $\log(1 + x)$. This normalisation allows comparison of expression levels across cells with different sequencing depths.

Normalisation was performed within each dataset before disease-level summaries were calculated. This ensured that downstream comparisons reflected the relative abundance of transcripts within each cell rather than differences in total sequencing depth between cells. The subsequent disease-level mean TP10K values should therefore be interpreted as expression summaries suitable for filtering and comparison, not as absolute molecular copy numbers. Because scRNA-seq suffers from dropout and variable capture efficiency, low but biologically meaningful expression may be underestimated. For this reason, the primary threshold was deliberately permissive. To test whether the conclusions depended on this permissive cutoff, networks were rebuilt at 0.01 and 0.05 TP10K and the main redundancy and robustness analyses were repeated on those higher-threshold networks.

For the spatial scRNA-seq datasets, transcript counts were normalised to 10 000 transcripts per cell and log1p transformed in the same way as for the non-spatial scRNA-seq datasets. This common transformation placed the spatial and non-spatial expression summaries on the same TP10K scale.

2.4 Cell type classification

Log1p-transformed data were analysed using CellTypist, a machine learning classifier for automated cell type annotation[139]. The Immune_All_High.pkl model was used for classification. CellTypist parameters were set with majority_voting = True and

mode = "prob_match". Cells were assigned as immune or non-immune based on the CellTypist output.

For visualisation, principal component analysis was performed on the CellTypist output using scanpy to identify 30 components that explained the most variation. Uniform Manifold Approximation and Projection was used for visualisation in two dimensions[138].

For the spatial scRNA-seq datasets, the same immune and non-immune classification framework was applied after expression matrices had been assembled with cell metadata. Non-immune cells were collapsed to a single non-immune category. All cells from the spatial scRNA-seq datasets were used when assessing ligand expression, and immune-classified cells were used when assessing receptor expression in the spatial networks.

2.5 Identification of cytokine and receptor genes

Human cytokine and receptor genes were identified by querying UniProt using a custom R script. The search terms used were "intercrine", "tumor-necrosis-factor-family", and "tumor-necrosis-factor-receptor-superfamily". Gene ontology terms queried were GO:0006935 for chemotaxis, GO:0004950 for chemokine receptor activity, GO:0005132 for interferon-alpha/beta receptor binding, GO:0005133 for interferon-gamma receptor binding, GO:0034348 for positive regulation of glial cell apoptotic process, GO:0004905 for type I interferon receptor activity, GO:0004906 for type II interferon receptor activity, GO:0005125 for cytokine activity, and GO:0004896 for cytokine receptor activity. Additional UniProt accessions O00155, Q16570, O95800, and P49685 were included.

2.6 Disease-level expression quantification

For each gene, the mean expression was calculated as mean transcript counts per 10 000 transcripts across all cells in the pooled samples for cytokine ligands, and across all immune cells for cytokine receptors. This aggregation of cell-level measurements into disease-level pseudo-bulk profiles provides a representative per-disease expression estimate.

Genes were considered expressed if mean TP10K exceeded 0.005, equivalent to approximately 0.5 transcript counts per million. Analyses were also performed using thresholds of 0.01 and 0.05 transcripts per 10,000 to ensure that redundancy and robustness results were not affected by choice of threshold.

The 0.005 TP10K threshold was used as the primary inclusion rule because the downstream question concerned whether a ligand or receptor was present in the disease-level repertoire, not whether it was among the most abundant transcripts. A very high cutoff would remove low-abundance receptors that may still be biologically relevant, particularly in heterogeneous immune populations where a receptor is expressed by a minority cell type. Conversely, a threshold of zero would retain transcripts likely to reflect noise or ambient RNA. The sensitivity analyses at 0.01 and 0.05 therefore tested whether the main conclusions depended on this permissive boundary.

2.7 Network construction

2.7.1 Ligand-receptor interaction data

Network edges were identified using ligand-target interaction data from the IUPHAR/BPS Guide to PHARMACOLOGY database[76]. The tsv files interactions.tsv and de-

tailed_endogenous_ligands.tsv were downloaded from `https://www.guidetopharmacology.org/download.jsp` and ligand-target information was merged and annotated with gene symbols. Edges recorded as antagonist in IUPHAR were excluded from the analysis because they do not represent receptor activation. Including antagonist relationships in the same unweighted graph as activating relationships would have counted inhibition as preserved signalling and would therefore have confounded the redundancy and robustness analyses.

2.7.2 Network representation

Networks were constructed as bipartite graphs with cytokine ligands and receptors as nodes and ligand-target interactions as connecting edges. For each disease and signalling category, expressed ligands were identified from all tissue cells and expressed receptors were identified from immune cells only. The IUPHAR interaction data were then used to identify edges between expressed ligands and receptors. Bipartite networks were constrained to contain no isolated ligands or receptors.

Networks were depicted as chord diagrams using Python. Separation into CK:CKR, NF-κB, and STAT networks, and clustering of ligand and receptor families, was based on published literature[14,58–60]. This resulted in 36 bipartite ligand-receptor networks comprising 12 diseases and 3 signalling mechanisms.

The comparison across the CK:CKR, NF-κB, and STAT axes was made at the level of network topology rather than biochemical mechanism. Each axis was reduced to the same representation: expressed ligands on one side of a bipartite graph, expressed immune-cell receptors on the other side, and curated activating ligand-receptor interactions as edges. This common abstraction allowed the same graph metrics to be applied to all three systems. It does not imply that chemokine receptors, TNF-family receptors, interleukin receptors, or interferon receptors signal through

identical biochemical mechanisms. The purpose of the abstraction was to compare how much redundant ligand-receptor connectivity each axis realises in diseased tissue under identical inclusion rules.

Only activating ligand-receptor relationships were included in the main cytokine networks. Antagonistic or decoy relationships were excluded because the redundancy and robustness analyses were designed to test the persistence of receptor activation under ligand loss. Including inhibitory edges in the same unweighted graph would have combined qualitatively different biological effects. This simplification means that the networks represent potential activating communication rather than the full regulatory logic of each signalling axis.

For the spatial scRNA-seq analysis, tiled local spatial axis networks were constructed using the same CK:CKR, NF-κB, and STAT interaction definitions used for the non-spatial cytokine-receptor networks. These definitions were mapped to genes measured in the CosMx 6K panels. Ligand-receptor pairs were included only when the required ligand and receptor genes were present in the panel and passed the same disease-level expression threshold. For multi-gene ligand or receptor entries, all component genes were required to satisfy the panel and expression rules.

Tiled spatial networks added a local tile requirement to the disease-level network definition. Local tile networks were constructed by dividing the tissue coordinate plane into square tiles. Within each tile, a ligand was counted as present when at least one cell had non-zero counts for the ligand genes. A receptor was counted as present when at least one immune cell had non-zero counts for the receptor genes. For multi-gene ligand entries, all ligand genes had to be detected in the same cell. For multi-gene receptor entries, all receptor genes had to be detected in the same immune cell.

Each CK:CKR, NF-κB, and STAT ligand-receptor interaction that passed the panel and disease-level expression filters was then tested in each square tile. A local edge was kept in a tile when at least one cell in that tile expressed the ligand and at least one immune cell in that tile expressed the receptor. The primary analysis used 500 μm square tiles and the 0.005 TP10K expression threshold. Tile-size sensitivity was assessed across 50–1500 μm square-tile side lengths in 50 μm increments, and transcript-abundance sensitivity was assessed at TP10K thresholds of 0.005, 0.01, and 0.05.

2.8 Network redundancy analysis

For each network, E denotes the complete set of observed ligand-receptor edges and $E_{\min}$ denotes a minimum edge cover, the smallest edge set that covers every node at least once. The edge counts are written as $|E|$ and $|E_{\min}|$, respectively. The redundancy index R, which measures the fraction of observed interactions beyond the minimum edge cover, was defined as:

$$R = 1 - \frac{|E_{\min}|}{|E|} \tag{2.1}$$

The minimum edge-cover count was calculated as $|E_{\min}| = |V| - M$, where V is the complete node set, $|V|$ is its size, and M is the size of a maximum matching. This derives from the Gallai identity $|V| = |E_{\min}| + M$ being true for graphs without isolated vertices. The maximum matching size was calculated in Python, representing the maximum number of ligand-receptor pairs that can be chosen such that no ligand or receptor is used more than once.

R ranges from 0, indicating no redundancy where every edge is necessary, to values

approaching 1, indicating high redundancy where many edges are redundant. Mean ligand promiscuity was calculated as $|E|$ divided by the number of ligand nodes, and mean receptor promiscuity as $|E|$ divided by the number of receptor nodes. Overall promiscuity was calculated as $2|E|/|V|$, where $|V|$ is the total number of ligand and receptor nodes; this gives the mean number of partners per node across both partitions.

The redundancy index is a topological measure and should not be interpreted as proving that every non-minimal edge is physiologically dispensable. A ligand-receptor interaction may be topologically redundant while still differing in affinity, kinetics, cell-type distribution, spatial localisation, or downstream signalling bias. The metric instead asks a narrower question: how many alternative edges exist beyond the minimum required to connect every expressed ligand or receptor at least once? This question is directly relevant to therapeutic blockade because a network with many alternative edges can retain connectivity after individual interactions are removed.

2.9 Network robustness analysis

2.9.1 Secondary extinction analysis

A primary extinction was defined as the removal of a cytokine node and all of its associated edges from the network[124]. A secondary extinction occurred when, as a consequence of primary extinctions, a receptor node lost all remaining connections to cytokine nodes. Primary extinctions were applied in 250 random permutations of cytokine node removal, continuing until all receptor nodes were eliminated.

For each permutation, a secondary extinction curve was generated by plotting the percentage of receptor nodes remaining against the percentage of cytokine nodes removed[124]. Functional network robustness was quantified as the area under the

median extinction curve, representing the median fraction of receptor nodes persisting throughout the cytokine removal process[120].

Random removal was used to model generic loss of cytokine ligands rather than a specific therapeutic sequence. This choice avoids assuming in advance which cytokines would be targeted first and allows robustness to be summarised as a property of the network as a whole. Because removal order affects the curve for any single permutation, 250 independent permutations were used for each network and the median curve was analysed. A receptor was considered functionally retained if at least one activating ligand remained connected to it. This is a conservative graph-level definition of functional persistence and does not account for differences in ligand potency.

2.9.2 Largest connected component robustness analysis

To quantify largest connected component robustness, the Schneider method was adapted to measure the persistence of the largest connected component as ligand nodes were progressively removed[125]. Let Q be the number of ligand nodes removed, N_0 the initial total number of graph nodes, and $S(Q)$ the fraction of those original nodes remaining in the largest connected component after removal of Q ligands. This fraction was recorded as:

$$S(Q) = \frac{\text{LCC size after removal of } Q \text{ cytokine ligand nodes}}{N_0} \tag{2.2}$$

Let L be the initial total number of ligand nodes. Largest connected component robustness R was computed as the mean of $S(Q)$ over the L removal steps:

$$R = \frac{1}{L} \sum_{Q=1}^{L} S(Q) \tag{2.3}$$

This measure captures the rate at which the largest connected component shrinks under ligand loss and is directly comparable across networks of different sizes because it is normalised to the initial node count[125].

Secondary extinction and largest connected component robustness capture complementary properties of the same ligand-removal simulation. Secondary extinction is calculated from receptors that retain at least one ligand after each removal step. Largest connected component robustness is calculated from the size of the largest connected component after each removal step, normalised to the initial node count. A network may preserve many individual receptor connections while the remaining graph is split into separate components, or it may retain a large connected component while specific receptors lose all ligands. Reporting both metrics therefore avoids relying on a single definition of robustness. For tiled spatial networks, secondary extinction AUC and largest connected component AUC were calculated using the same 250 random ligand-removal simulations and summary definitions as the main non-spatial network analysis.

2.10 Identification of chemokine interactions

Chemokine:receptor activating interactions were identified from IUPHAR-DB as described above, accessed in November 2025[61]. Chemokine:chemokine interactions were identified using PSICQUIC queries to the interaction providers BioGrid, IntAct, IMEx, MINT, and UniProt. Physical interactions were restricted to those supported by NMR, X-ray crystallography, biolayer interferometry, or surface plasmon resonance experiments.

Chemokine:evasin interactions were identified from the literature[6,126,131–133,155]. This resulted in 386 evasin:chemokine interactions, 71 chemokine:chemokine interactions,

and 66 activating chemokine:receptor interactions.

2.11 Structural prediction with AlphaFold 3

Protein sequences for chemokines and dimeric partners were obtained from UniProt. The signal peptide was removed where present because secreted chemokines are cleaved to mature extracellular proteins before binding their partners. Removing this segment ensured that AlphaFold 3 modelled the mature binding sequence rather than an N-terminal peptide that is not present in the functional secreted protein. Protein sequences were prepared into JSON files and submitted to the AlphaFold 3 webserver[146]. The output was downloaded and the highest-ranking model was analysed.

For comparison of AlphaFold 3 models with published structures, all PDB structures related to human chemokines were identified from UniProt. Chemokine:receptor, chemokine:evasin, and chemokine:chemokine X-ray and EM structures were identified. Where alternative structures were available, the structure with the highest resolution was chosen.

AlphaFold 3 predictions were used as structural hypotheses for systematic comparison rather than as substitutes for experimental structures. The validation step against available PDB complexes was therefore essential for assessing whether the predicted binding poses were plausible in interaction classes represented in structural databases. For complexes without experimental structures, the same modelling and analysis rules were applied, but conclusions were interpreted at the level of predicted interface geometry.

2.12 Structural analysis methods

2.12.1 RMSD calculation

Root Mean Square Deviation, abbreviated RMSD, was calculated using PyMOL v2.5.7. RMSD measures the average distance between equivalent atoms after two structures have been aligned, so lower values indicate closer structural agreement. For PDB structures with greater than two chains, the lowest calculated RMSD between the AlphaFold 3 dimeric complex and any pair of chains from the PDB structure was reported.

2.12.2 Analysis of dimeric partner binding sites on chemokines

Interchain residue contacts within 5 Å of a dimer partner were identified using PyMOL v2.5.7 and used to define binding surfaces on the chemokine. Multiple sequence alignments were produced using Clustal Omega v1.2.3. For each MSA residue position, the fraction of complexes contacted by a receptor or by an evasin was calculated. A threshold of 0.3 means that a residue position was classified as an interaction hotspot if it was contacted in at least 30% of modelled complexes in that interaction class.

2.12.3 Binding site overlap analysis

To quantify overlap between two binding-site residue sets, denoted A and B, the Jaccard index $J(A, B)$ was computed[156]:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.4}$$

where $|A \cap B|$ is the number of residues shared by the sets and $|A \cup B|$ is the number present in either set. Statistical significance of residue-level overlap was assessed

using a permutation test for the Jaccard index. For each pairwise comparison, one contact vector was held fixed while the other was randomly permuted 3 000 times. The empirical p-value was calculated as the proportion of random permutations yielding a Jaccard index equal to or greater than the observed value. This non-parametric approach avoids the independence assumptions inherent to Fisher's exact or chi-squared tests and remains valid when neighbouring residues are spatially or sequentially correlated.

2.12.4 Analysis of chemokine binding sites on evasins

Contacts between each evasin and its chemokine partner were identified in PyMOL v2.5.7 using a 5 Å interchain distance cutoff. The resulting evasin surfaces were encoded as binary contact vectors for pairwise overlap analysis.

2.12.5 Conformational plasticity analysis

To analyse conformational plasticity at evasin binding interfaces, the template modelling score was calculated using US-Align following alignment of non-interface and interface residues[157]. The TM-score ranges from 0 to 1, where higher values indicate greater similarity and thus lower plasticity.

For each evasin with at least two bound chemokines, pairwise TM-score matrices were computed for non-interface regions and interfaces. Each matrix was summarised by its median pairwise TM-score.

2.12.6 Interface residue conservation

For two interface residue sets, X for the first interface and Y for the second, the Sørensen-Dice coefficient (SDC) was calculated as[158]:

$$\mathrm{SDC} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{2.5}$$

The numerator counts shared residue indices twice, while the denominator is the sum of the two set sizes. Evasin-level summaries were the median pairwise coefficients.

2.13 Statistical analysis

Statistical analyses were performed using Python, including scipy. Statistical tests used are described in figure legends. Data are displayed as boxplots with an inner horizontal line indicating the median and whiskers representing the 25th and 75th percentiles, or as violins with horizontal lines marking the median, 25th percentile, and 75th percentile.

2.14 Software used

Most analyses were performed in Python. Python packages used included celltypist 1.6.2, scanpy 1.9.6, anndata 0.10.4, pandas 2.2.3, numpy 1.26.4, seaborn 0.13.2, matplotlib 3.7.1, scipy 1.13.0, and pymol 2.5.7. R was used only for the UniProt cytokine and gene-name extraction step; R version 4.5.0 was used with packages including AnnotationDbi_1.71.0, Biobase_2.69.0, BiocGenerics_0.55.0, biomaRt_0.65.0, Biostrings_2.77.2, and dplyr_1.1.4. US-Align was obtained from the Zhang group website[157].

Chapter 3

Analysis of Single-Cell RNA-Sequencing Data from Twelve Immunoinflammatory Diseases

3.1 Introduction

The failure of chemokine-targeted therapies in immune-mediated inflammatory diseases remains one of the most significant unresolved problems in immunopharmacology[7,8,45]. TNF inhibitors, IL-6 receptor antagonists, and JAK inhibitors have achieved clinical success in rheumatoid arthritis, inflammatory bowel disease, and psoriasis[9–11]. Yet despite three decades of clinical trials, no chemokine receptor antagonist has achieved approval for any inflammatory indication[7,8,100]. This disparity is puzzling because chemokines direct the leukocyte infiltration that defines these diseases[57,58]. Understanding the molecular basis of this therapeutic failure requires systematic characterisation of cytokine signalling as it operates in human disease tissues.

Cytokine signalling in inflammation can be organised into three major axes based on downstream effector mechanisms[11,14]. Chemokines constitute the largest cytokine family, with 46 ligands signalling through 18 G protein-coupled receptors to direct

immune cell trafficking[58,61]. The NF-κB axis comprises TNF, IL-1, and IL-17 family cytokines, which converge on NF-κB nuclear translocation and pro-inflammatory gene expression[14,59]. The STAT axis includes interferons and the IL-2, IL-6, IL-10, and IL-12 families, which signal via Janus kinases to activate STAT transcription factors[14,60]. These axes are interconnected: NF-κB and STAT pathways induce chemokine gene transcription, while chemokine-driven recruitment of cytokine-producing cells amplifies local inflammation through feed-forward loops[63–65].

Single-cell RNA sequencing enables characterisation of immune responses in human disease at cellular resolution[135]. Bulk RNA sequencing averages gene expression across thousands of cells, obscuring the cellular heterogeneity that defines inflamed tissues[135]. In contrast, scRNA-seq captures the transcriptome of individual cells, revealing subpopulations with distinct functional states and cytokine repertoires[135,139]. This resolution enables identification of which cell types produce specific cytokines and which immune cells express their cognate receptors, allowing inference of communication pathways that operate in vivo[135].

Previous scRNA-seq studies have characterised immune infiltrates in individual diseases, including atherosclerosis[151], type 1 diabetes[25], inflammatory bowel disease[152], and rheumatoid arthritis[46]. However, no study has systematically compared cytokine expression patterns across multiple diseases using a standardised analytical framework. Such a comparison is essential for distinguishing disease-specific features from conserved properties of inflammatory signalling. Moreover, previous work has not explicitly compared expression patterns across the three cytokine axes, which is necessary for understanding why therapeutics targeting the NF-κB and STAT axes succeed where chemokine-targeted approaches fail.

This chapter describes the analysis of scRNA-seq datasets from twelve immune-

mediated inflammatory diseases using a standardised computational pipeline. I characterise the immune cell populations present in inflamed tissues, quantify cytokine ligand and receptor expression across the three signalling axes, and examine the relationship between ligand and receptor diversity. I then use two public spatial scRNA-seq datasets generated with the CosMx 6K platform, one from psoriasis skin and one from ulcerative colitis rectal tissue. These datasets report gene expression for individual cells, record where each cell was located in the tissue, and include enough chemokine, NF-κB, and STAT ligand and receptor genes for the analysis. The goal is to establish the molecular foundation for understanding why chemokine networks may differ fundamentally from NF-κB and STAT networks in inflammatory disease. This foundation can then be built upon to interrogate differences between these signalling pathways that could explain why drugs targeting chemokines have failed while those targeting the NF-κB or STAT axes have succeeded.

3.2 Results

3.2.1 ScRNA-seq data from twelve immune-mediated inflammatory diseases were selected for analysis

ScRNA-seq datasets from twelve immune-mediated inflammatory diseases were analysed. The conditions covered several organ systems, contributed substantially to global disease burden, and met the inclusion criteria in **Table 2.1**. Datasets were chosen so that cells had not been physically pre-sorted, to avoid cell population bias.

The diseases included atherosclerosis[151], type 1 diabetes[25], Crohn's disease and ulcerative colitis[152], idiopathic pulmonary fibrosis[153], psoriasis[154], systemic lupus erythematosus and discoid lupus erythematosus[39], Sjögren's syndrome[42], rheumatoid arthritis and

psoriatic arthritis[46], and endometriosis[50]. No chemokine-targeted drug is approved for an inflammatory indication in any of these conditions, whereas several conditions have approved therapies that target NF-κB- or STAT-linked pathways[7–10,100].

The selection was designed to test whether inflammatory signalling features recur across different anatomical settings rather than within a single organ system. The dataset collection therefore included vascular tissue, pancreatic islets, intestinal mucosa, lung, skin, salivary gland, synovium, and ectopic endometrium[25,39,42,46,50,151–154]. These tissues differ in resident stromal cells, barrier exposure, baseline immune surveillance, and disease-specific histology[25,39,42,46,50,151–154]. If a signalling feature is observed across this set, it is less likely to reflect a peculiarity of one tissue type. Conversely, differences between diseases remain informative because they show how a shared inflammatory architecture is modified by local tissue context.

This design also separates two questions. The first is whether different diseases contain similar immune cell classes and cytokine axes, which can be addressed using single-cell expression and cell-type annotation. The second is whether the ligand-receptor networks formed by those axes have different topological properties, which requires the graph analysis in the following chapter. Chapter 3 addresses the first question and establishes the foundation for the second.

3.2.2 A computational pipeline was constructed to process these datasets uniformly

A standardised computational pipeline was developed to analyse all scRNA-seq datasets uniformly. Substantial time was required to adapt the pipeline to heterogeneous public submissions, particularly GEO records, because file naming, matrix layout, feature annotation, and metadata structure varied between source datasets. The pipeline

identified cytokine ligands produced by all tissue cell types and cytokine receptors present on immune cells. Raw transcript abundances were extracted, quality-controlled, normalised to 10 000 transcripts per cell, and classified using CellTypist[139]. The same pipeline was applied to all datasets to allow cross-disease comparison; ligands were quantified across all tissue cell types and receptors on immune cells.

This ligand-versus-receptor split reflects the biology of inflammatory communication. Cytokines and chemokines can be produced by immune cells, stromal cells, epithelial cells, endothelial cells, and other tissue-resident populations. Receptor expression relevant to leukocyte recruitment and/or activation, however, was assessed on immune cells because the central therapeutic question concerns signalling into immune populations that drive inflammation. The resulting expression summaries therefore describe tissue-level cytokine supply and immune-cell receptor availability within the same disease sample.

3.2.3 CellTypist classified cells as immune or non-immune

Immune and non-immune cells were identified using CellTypist, a machine-learning classifier[139]. Expression of conventional immune cell markers was higher in the immune cell group than in the non-immune group. The immune cell types identified included T-cells, B-cells, macrophages, monocytes, dendritic cells, plasmacytoid dendritic cells, innate lymphoid cells, plasma cells, and mast cells. No neutrophils were identified in this analysis; an explanation for this is offered in the Discussion.

3.2.4 Immune cell distribution was visualised in heatmaps and UMAP

The distribution of immune and non-immune cell subtypes across inflamed tissues was visualised in four complementary ways (**Figures 3.1 to 3.4**). First, a heatmap of conventional immune cell marker expression compared immune and non-immune subsets per disease, with samples grouped or labelled by compartment, e.g. immune [_i] vs. non-immune [_ni], on the horizontal axis and marker genes on the vertical axis (**Figure 3.1**).

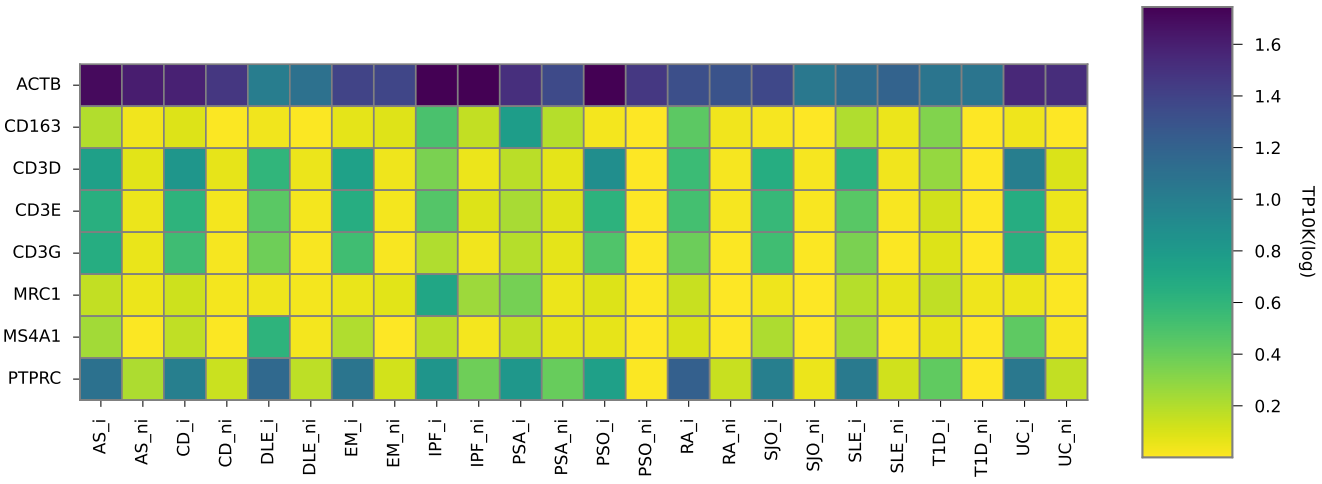

Figure 3.1: Immune marker gene expression in diseased tissue compartments across the twelve immune-mediated inflammatory diseases. Rows show selected marker genes and columns show disease-level cell compartments, with _i denoting immune cells and _ni denoting non-immune cells. Values represent mean expression after TP10K normalisation and $\log_{10}(\text{TP10K} + 1)$ transformation. Higher colour intensity indicates higher marker expression, with values capped at the first and 99th percentiles for visualisation.

Second, a clustered heatmap showed the proportion of each immune and non-immune cell subtype across the twelve diseases, giving a direct view of how cell type composition varied by condition (**Figure 3.2**). Rows represent cell types and columns represent diseases.

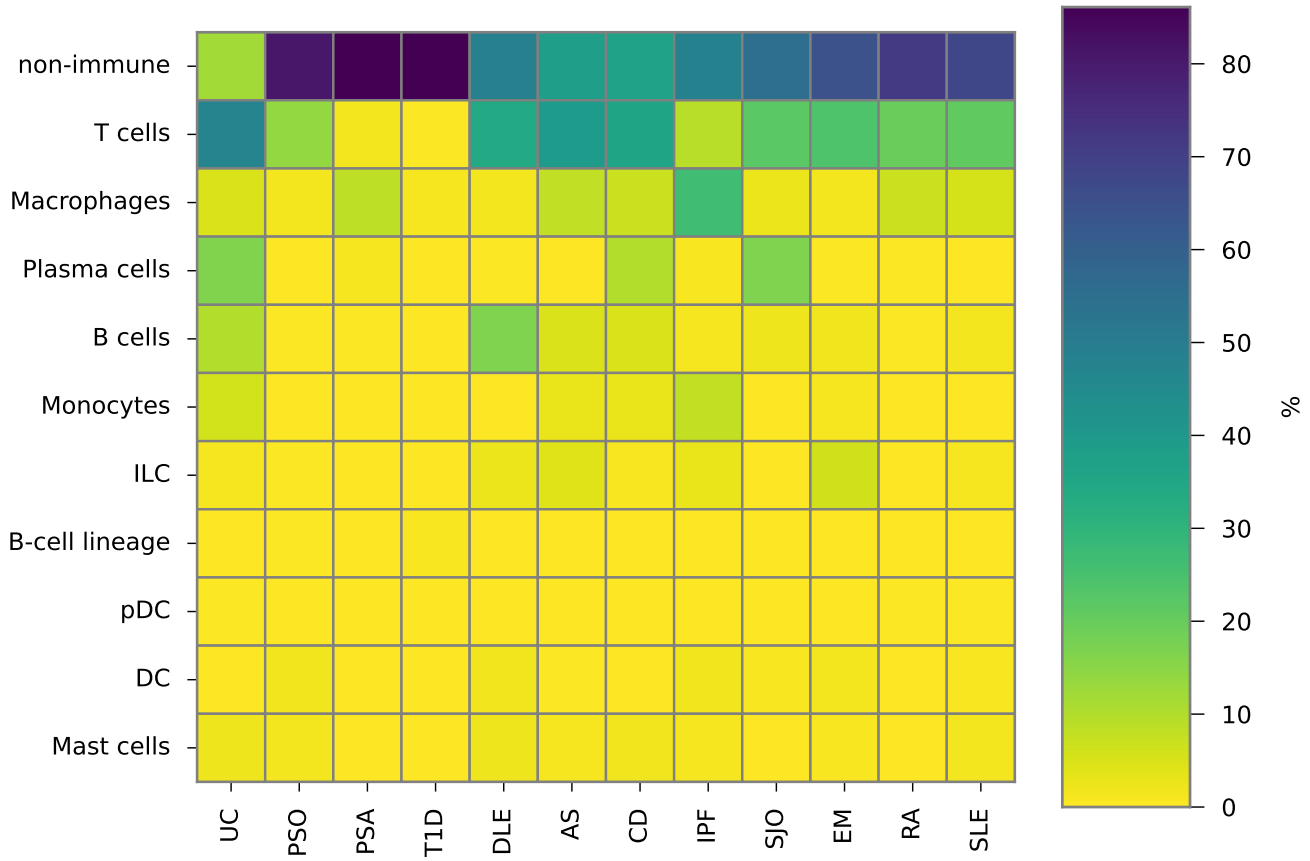

Figure 3.2: Cell type composition across the twelve immune-mediated inflammatory diseases. Rows show CellTypist cell-type annotations and columns show disease abbreviations. Values represent the mean percentage of cells assigned to each cell type within each disease. Rows and columns were clustered by similarity in cell-type abundance profiles, and values were capped at the first and 99th percentiles before plotting.

Third, Uniform Manifold Approximation and Projection (UMAP)[138] was applied to the single-cell transcriptomes; each point is coloured by its CellTypist-assigned cell type, so that clustering and overlap between populations can be seen (**Figure 3.3**). This projection allows cell-type populations and their overlap to be compared across diseases.

Fourth, cell type abundances were summarised as boxplots across the twelve diseases (**Figure 3.4**). The plots show the distribution of disease-level abundance values for

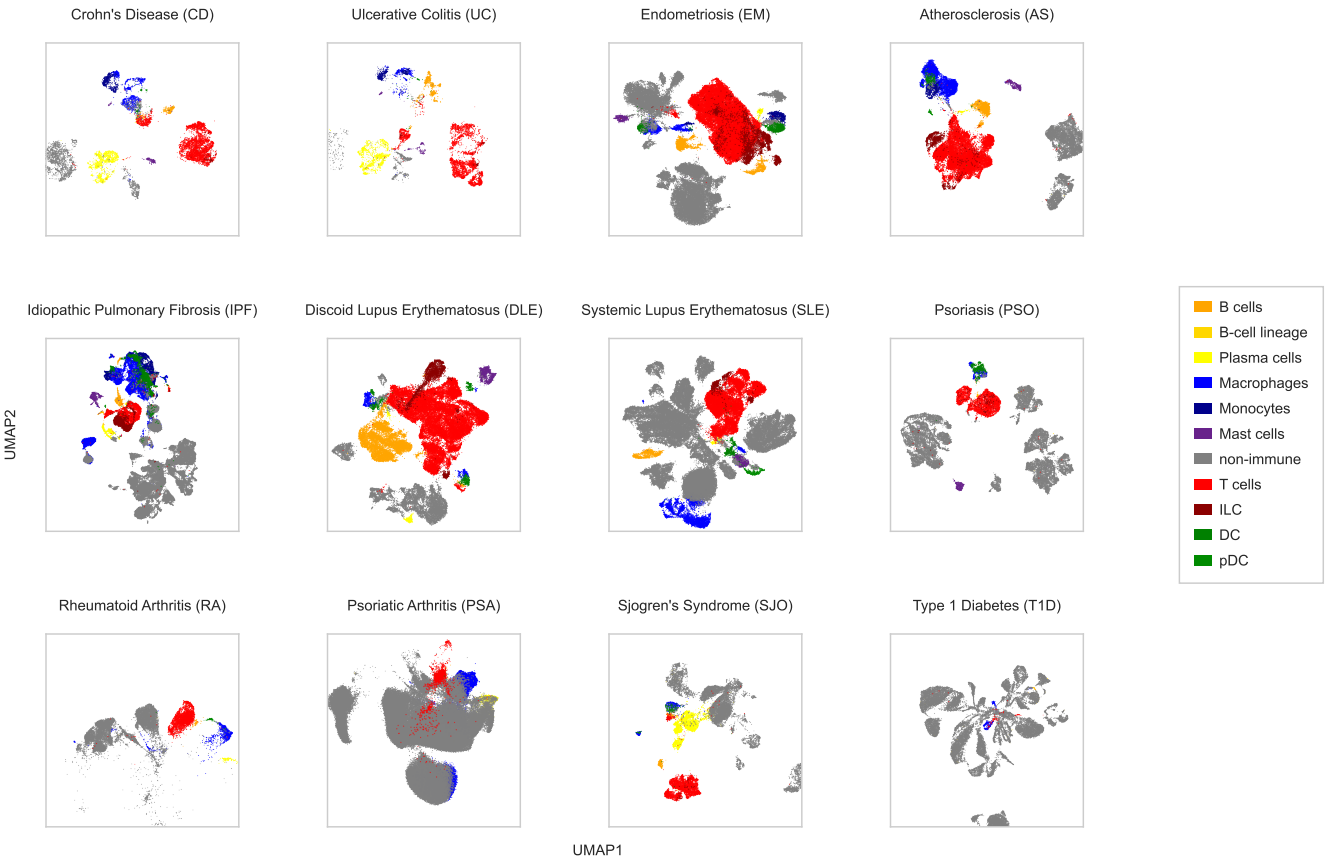

Figure 3.3: Uniform Manifold Approximation and Projection (UMAP) visualisation of cells from diseased samples across the twelve immune-mediated inflammatory diseases. Each panel shows one disease, and each point represents a single cell positioned by its UMAP1 and UMAP2 coordinates after preprocessing and CellTypist annotation. Points are coloured by final CellTypist cell-type prediction, with non-immune cells collapsed into a single grey category and immune populations shown by distinct colours.

each cell type.

Taken together, these visualisations showed that immune and non-immune compartments separated by conventional marker expression, supporting the CellTypist-based classification. T cells and macrophages together comprised most of the immune compartment in every disease, ranging from 54.6% in Sjögren's syndrome to 93.7% in rheumatoid arthritis.

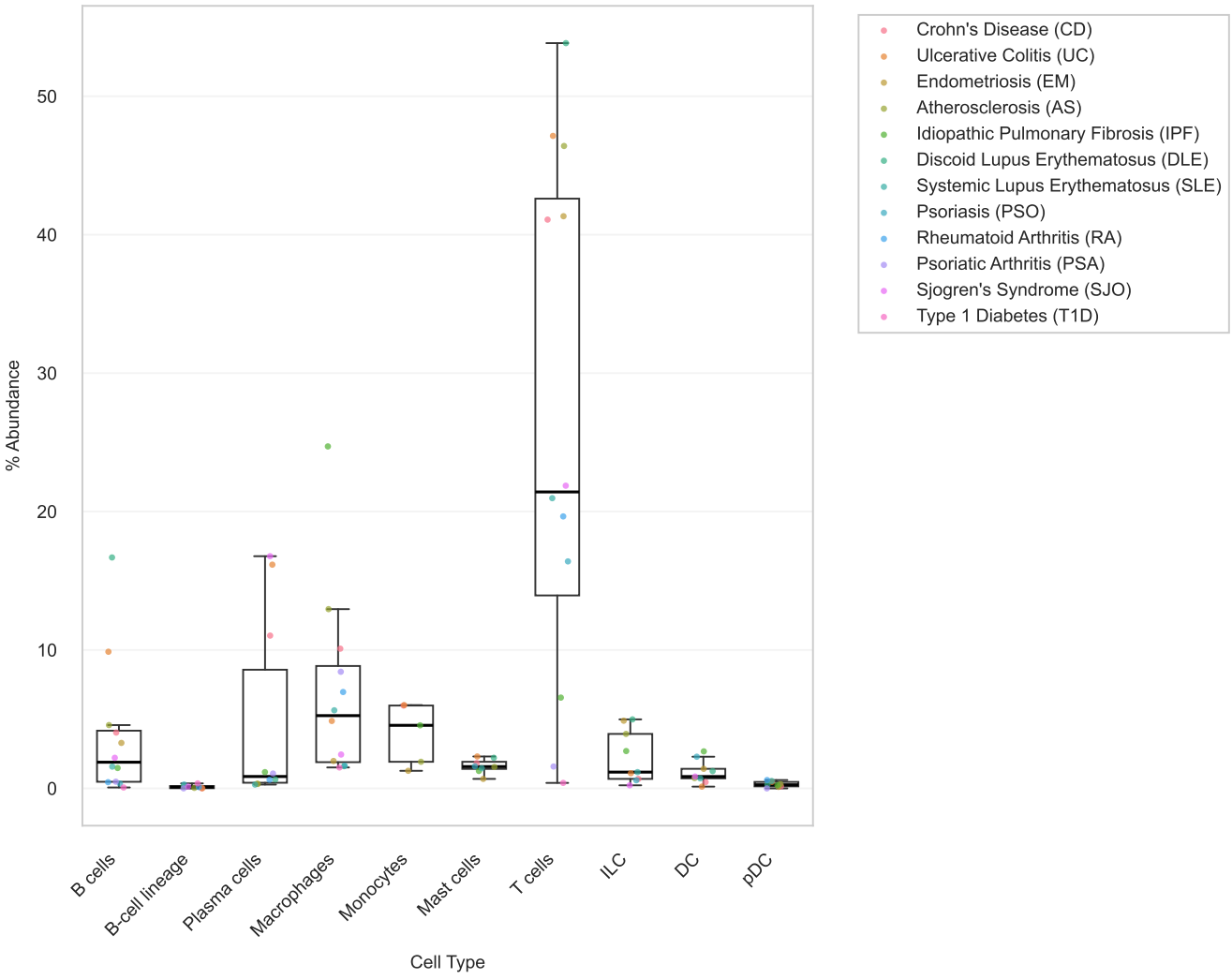

Figure 3.4: Cell type abundance distributions across the twelve immune-mediated inflammatory diseases. The x-axis shows CellTypist cell-type categories and the y-axis shows percentage abundance within each disease dataset. Boxes summarise the distribution across diseases for each cell type, with the central line showing the median and the box spanning the interquartile range. Whiskers extend to the most extreme values within 1.5 times the interquartile range. Coloured points show individual disease-level values, with colours corresponding to the disease legend.

T cells were the dominant immune class in most datasets, representing 74.7% of immune cells in endometriosis, 70.9% in psoriasis, 69.2% in rheumatoid arthritis, 66.1% in DLE, 64.8% in SLE, 63.8% in atherosclerosis, 54.5% in Crohn's disease, and 53.5% in ulcerative colitis. Macrophages were the largest immune class in IPF and

psoriatic arthritis, representing 54.3% and 72.8% of immune cells, respectively. Plasma cells were prominent in Sjögren's syndrome, ulcerative colitis, and Crohn's disease, representing 37.7%, 18.4%, and 14.6% of immune cells, respectively. B cells were most prominent in DLE and ulcerative colitis, representing 20.5% and 11.2% of immune cells, respectively.

The principal between-disease differences therefore lay in the balance between lymphoid and myeloid compartments and in the abundance of smaller populations such as B cells, dendritic cells, plasma cells, mast cells, and innate lymphoid cells. This pattern indicated that each disease retained a distinct cellular composition despite a shared inflammatory architecture.

3.2.5 Fold change in cytokine and receptor expression was evaluated for nine diseases

Matched healthy controls were available within the same dataset for eight of the twelve diseases: discoid lupus erythematosus, systemic lupus erythematosus, psoriasis, type 1 diabetes, endometriosis, Crohn's disease, ulcerative colitis, and idiopathic pulmonary fibrosis. For atherosclerosis, no healthy donor control exists for plaque tissue; the comparison was therefore to proximal adjacent unaffected tissue from the same study. Accordingly, fold-change analysis was performed for nine diseases with either matched healthy or adjacent comparator tissue. The remaining three diseases were Sjögren's syndrome, rheumatoid arthritis, and psoriatic arthritis; these did not have healthy control samples in the same dataset. Genes were considered expressed when mean TP10K exceeded 0.005. For these nine comparator analyses, the fold change in ligands, immune cell receptors, and pathway target genes for the NF-κB and STAT axes was evaluated (**Figure 3.5**). Inflamed tissues showed a diversity of cytokine ligand and receptor expression patterns.

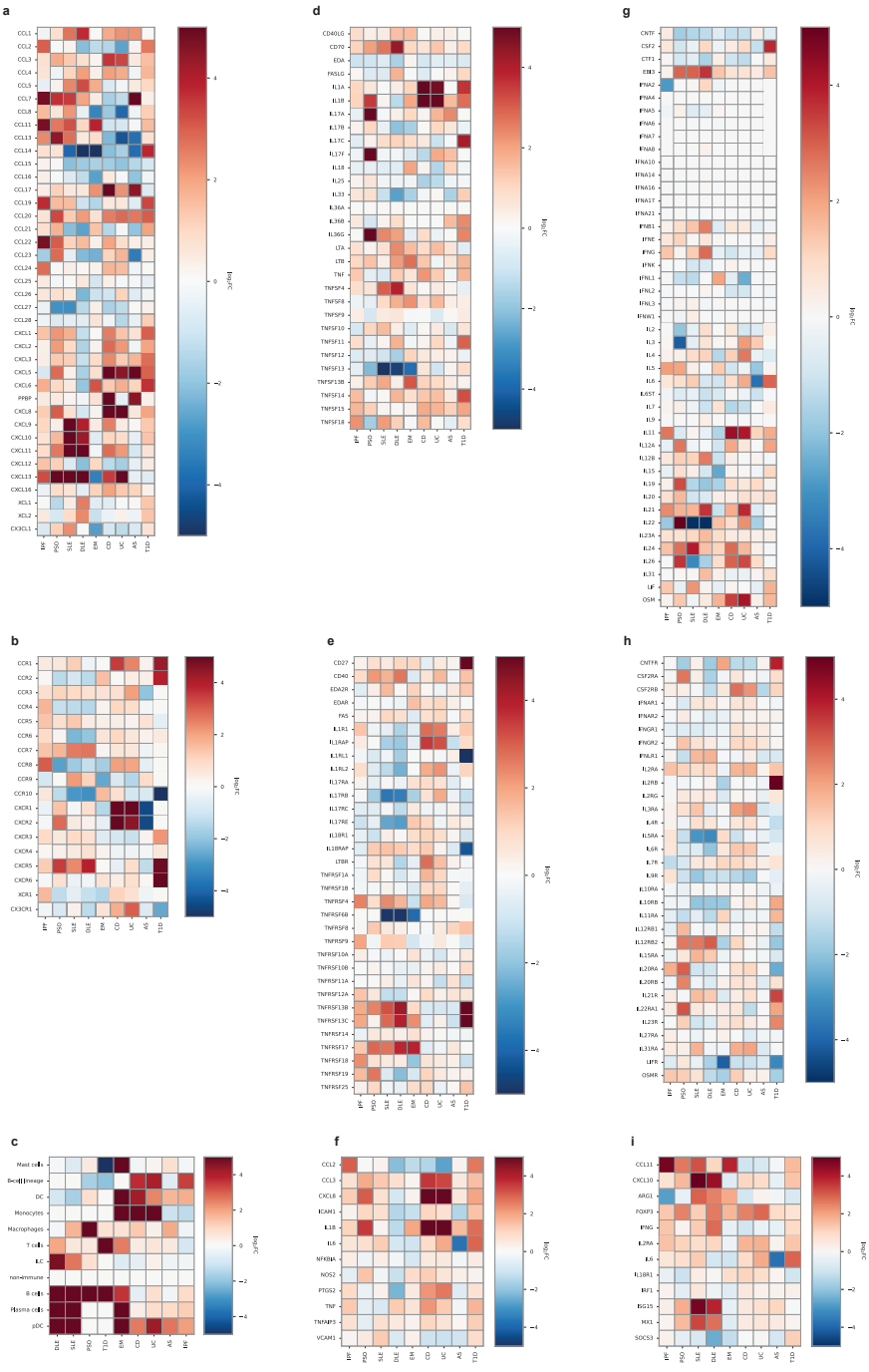

Figure 3.5: Disease-associated $\log_2$ fold change in inflamed tissue relative to matched healthy or adjacent comparator tissue. Panels **a**–**c** show CK:CKR ligands, receptors, and immune-cell abundance; panels **d**–**f** show NF-κB ligands, receptors, and target genes; and panels **g**–**i** show the corresponding STAT measurements. Rows show genes or cell types and columns show diseases. Red indicates increased expression or abundance, blue indicates decreased expression or abundance, and white indicates little or no change.

As an indirect functional readout consistent with chemokine pathway activation and chemotaxis, fold change in immune cell counts observed in the disease tissues was examined alongside direct measurements of chemokine ligand and chemokine receptor expression. Macrophages and monocytes, mast cells, dendritic cells and plasmacytoid dendritic cells, innate lymphoid cells, T-cells, B-cells, and plasma cells were enriched in inflamed tissues. Inflamed tissues also showed increased expression of downstream markers of NF-κB and STAT signalling. Together, these direct expression measurements and indirect recruitment-associated readouts showed substantial disease-specific variation across the three major signalling pathways. Expression levels (transcripts per 10k) for ligands and receptors across the CK:CKR, NF-κB, and STAT axes are shown in **Figure 3.6**.

The fold-change analysis had a different role from the later network analysis. Fold change identifies how disease tissue differs from matched healthy or adjacent comparator tissue where such tissue was available. It therefore provides evidence that inflammatory tissues are transcriptionally altered relative to comparator samples. The network analysis, by contrast, uses disease-state expression to ask which ligand-receptor connections are available within inflamed tissue. A gene can be important for network construction even if it is not strongly induced relative to control, provided that it is expressed in the diseased tissue.

IPF showed increased CK:CKR expression, consistent with a stronger chemokine-recruitment signal in that dataset. Crohn's disease and ulcerative colitis showed increased NF-κB receptor expression, consistent with prominent inflammatory cytokine responsiveness in intestinal inflammation. These disease-specific patterns indicate that the same signalling axes were not uniformly altered across conditions, but instead varied by tissue and disease context.

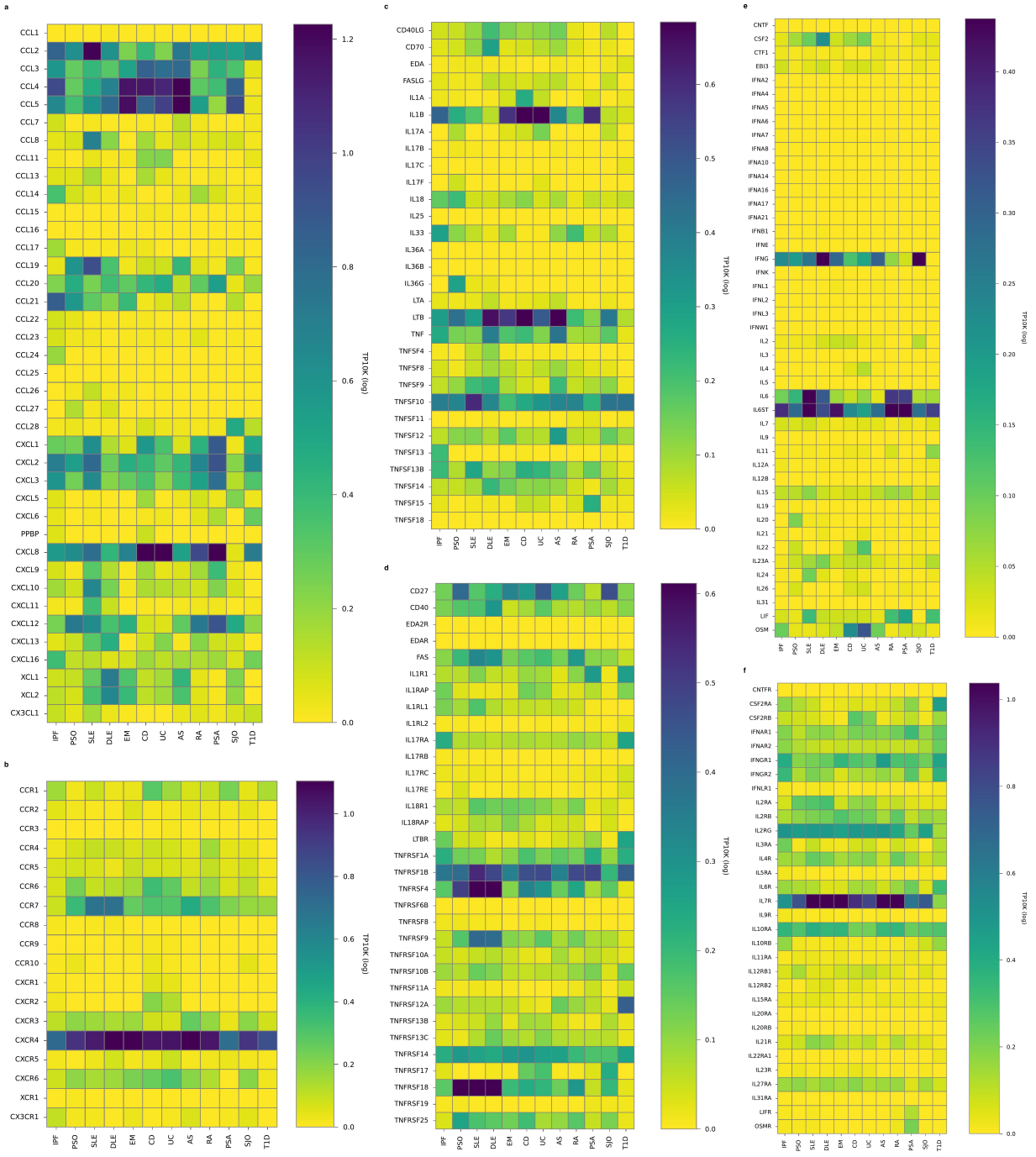

Figure 3.6: Mean disease-state expression after TP10K normalisation and $\log_{10}(\text{TP10K}+1)$ transformation across the twelve diseases. Panels **a**–**b**, **c**–**d**, and **e**–**f** show ligands and receptors for the CK:CKR, NF-κB, and STAT axes, respectively. Rows show genes and columns show disease abbreviations. Yellow indicates low expression and purple indicates high expression; values were capped at the first and 99th percentiles for visualisation.

3.2.6 Chemokine ligand-receptor asymmetry was conserved across diseases

Expression was next summarised as counts of expressed ligands and receptors per disease condition, using the same TP10K threshold of 0.005 applied throughout this thesis. For the CK:CKR axis, the number of expressed ligands exceeded the number of expressed receptors in all twelve of the twelve disease conditions examined. This pattern also held in the nine disease conditions for which matched healthy or adjacent comparator tissue was available.

Across the full twelve-disease set, the median ligand:receptor ratio was 2.38 for CK:CKR, compared with 0.74 for the NF-κB axis and 0.55 for the STAT axis. Restricting the analysis to the nine diseases with matched healthy or adjacent comparator tissue gave nearly identical results, with median ligand:receptor ratios of 2.31 for CK:CKR, 0.75 for NF-κB, and 0.56 for STAT. Thus, the chemokine axis consistently displayed ligand excess, whereas the NF-κB and STAT axes showed more balanced or receptor-heavy organisation.

These results identify a conserved transcriptional asymmetry in human inflammatory disease tissues. The CK:CKR axis is characterised by a larger repertoire of expressed ligands than receptors, and this asymmetry is robust both across all twelve disease conditions and within the subset with comparator tissue. This provides one aspect of the molecular context for the network analysis in the following chapter, where ligand excess is tested as a potential substrate for redundant connectivity.

The distinction between ligand excess and network redundancy should be clarified: redundancy depends on which ligands bind which receptors, not only on how many ligands are expressed. A system could contain many ligands but still be sparse if each ligand had a single private receptor. However, when ligand excess is combined with

known ligand-receptor promiscuity, it creates the conditions under which redundant connectivity can emerge. These non-spatial expression results therefore established the expression pattern that makes the formal graph analysis biologically plausible.

3.2.7 Spatial single-cell RNA-seq datasets

Public spatial scRNA-seq datasets were screened for two requirements. First, each dataset had to report gene expression for individual cells and record where each cell was located in the tissue. Second, the gene panel had to measure enough genes from the chemokine, NF-κB, and STAT ligand and receptor lists. The screen compared each panel with a list of 195 genes used in this thesis: 57 chemokine genes, 61 NF-κB genes, and 77 STAT genes. Two public spatial scRNA-seq datasets satisfied both requirements: GSE314158 from psoriasis skin and GSE312415 from ulcerative colitis rectal tissue. Both were generated with the CosMx 6K platform[149,150]. CosMx 6K is an imaging-based assay that measures a fixed panel of approximately 6 000 genes in individual cells and preserves each cell's position in the tissue[149,150]. The panel comparison is shown in **Figure 3.7**.

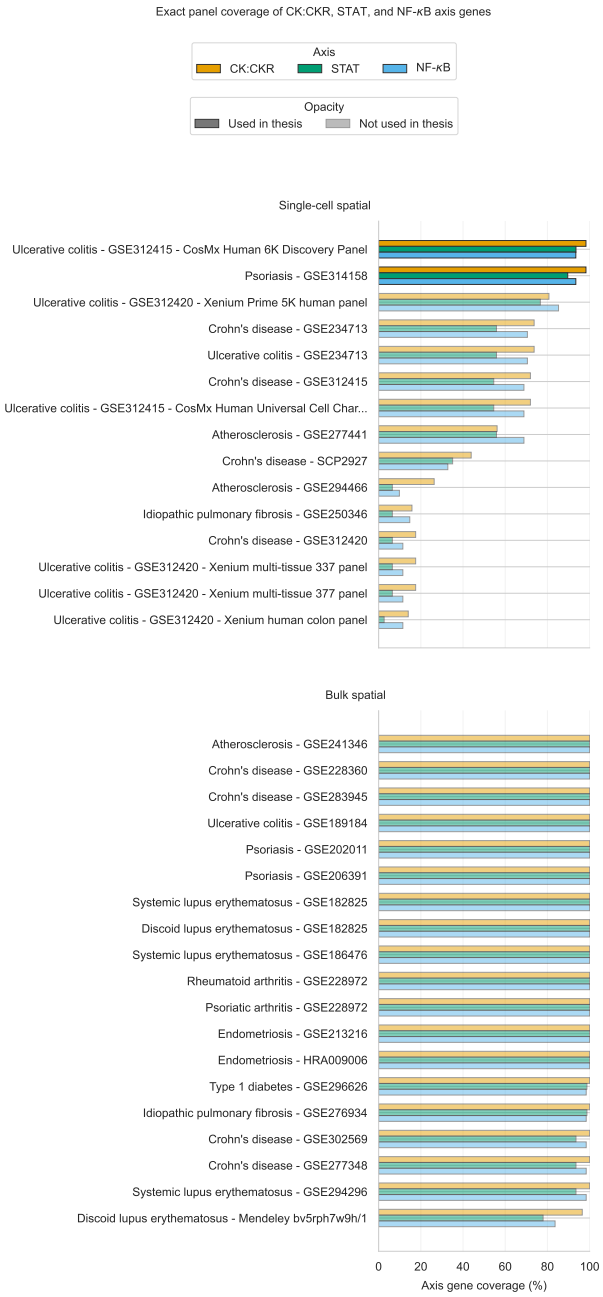

Figure 3.7: Coverage of chemokine, NF-κB, and STAT ligand and receptor genes in public spatial scRNA-seq datasets. Bars show how many genes from the thesis chemokine, NF-κB, and STAT ligand and receptor lists were measured by each assay panel. The full comparison list contained 195 genes. Datasets are grouped by spatial assay type. The psoriasis skin and ulcerative colitis rectal tissue datasets analysed in later sections are shown at full opacity, and other datasets are shown at 50% opacity.

The two public spatial scRNA-seq datasets used here reported gene expression for individual cells identified in the tissue images. The maps and percentage summaries showed that both datasets contained several annotated cell types rather than one predominant class (**Figures 3.8 and 3.9**).

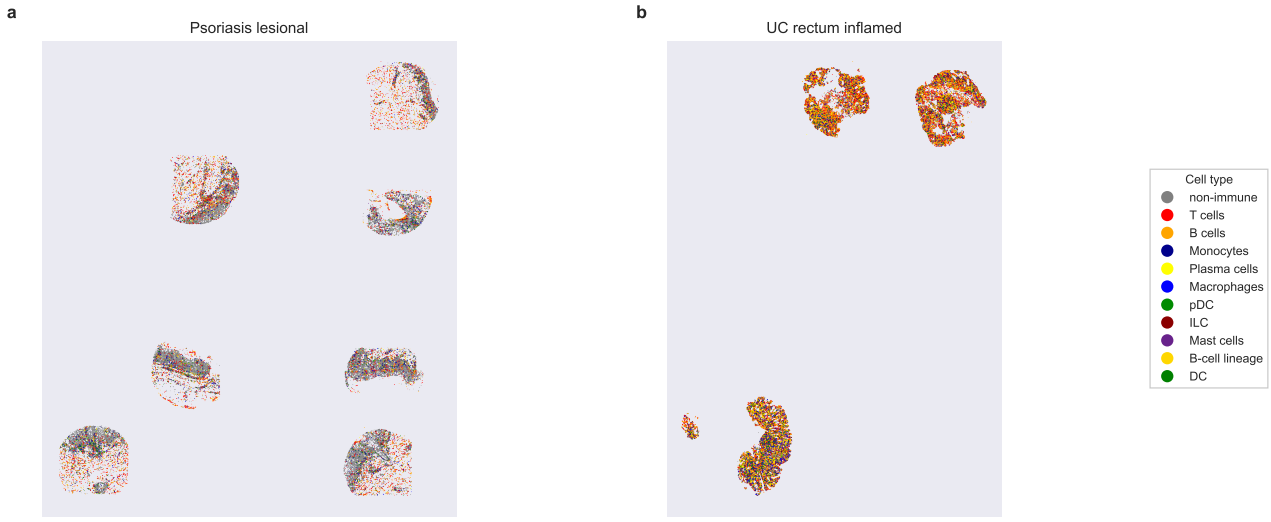

Figure 3.8: Maps of cells in the two public spatial scRNA-seq disease tissues used for the analysis. Panel **a** shows psoriasis lesional skin, and panel **b** shows ulcerative colitis inflamed rectal tissue. Each point is one cell at its measured position in the tissue. Colours show CellTypist cell-type annotation.

They also reported each cell's position, broad CellTypist labels, and whether each cell was classified as immune or non-immune. For psoriasis, lesional and non-lesional skin were used to describe expression patterns in this chapter, but only lesional skin was used for the local network analysis in the following chapter. Ulcerative colitis contributed inflamed rectal tissue.

Chemokine, NF-κB, and STAT ligands and receptors were detectable in the two public spatial scRNA-seq datasets used for the analysis. Ligands were summarised across all cells. Receptors were summarised across immune cells. The same rule was used for the non-spatial datasets. These results showed that both spatial scRNA-seq datasets measured the genes needed to build local networks in the following chapter, while

redundancy and robustness were evaluated only in that network analysis (**Figure 3.10**).

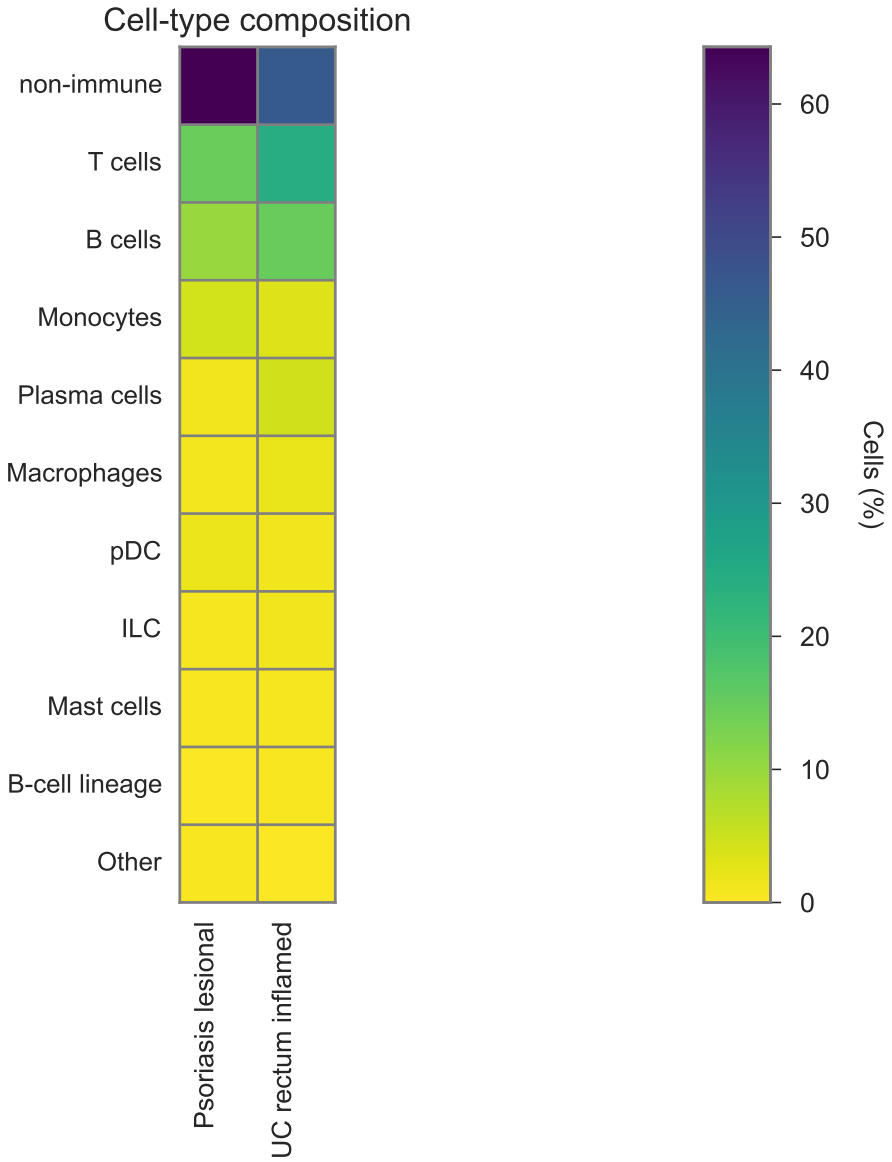

Figure 3.9: Cell-type composition of the two public spatial scRNA-seq disease tissues used for the analysis. Values show the percentage of cells assigned to each of the most abundant CellTypist cell-type categories, with less frequent categories grouped as Other.

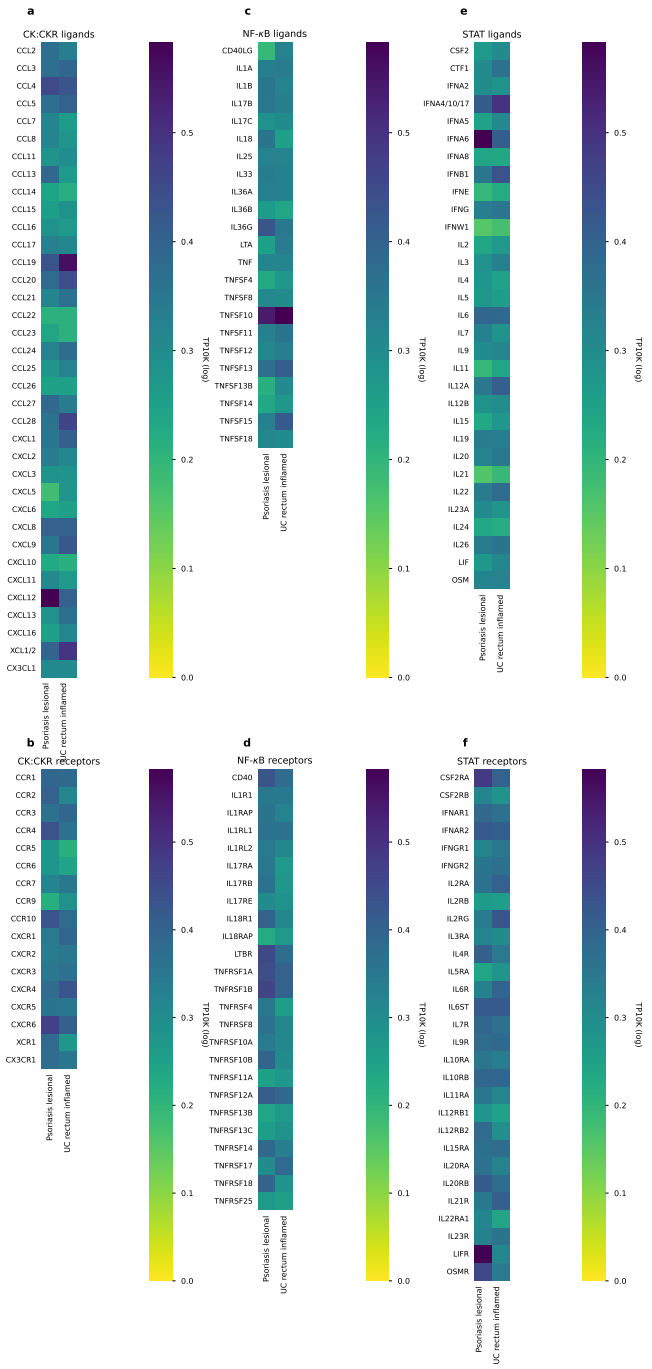

Figure 3.10: Chemokine, NF-κB, and STAT ligand and receptor expression in the two public spatial scRNA-seq disease tissues used for the analysis. Panel **a** shows chemokine ligands, panel **b** shows chemokine receptors, panel **c** shows NF-κB ligands, panel **d** shows NF-κB receptors, panel **e** shows STAT ligands, and panel **f** shows STAT receptors. Values show mean TP10K expression summaries. Ligands were summarised across all cells, and receptors were summarised across immune cells.

3.3 Discussion

The cell-type analysis identified T-cells, B-cells, monocytes/macrophages, and dendritic cells across the disease panel, although their distributions differed by disease (**Figures 3.2 and 3.4**). The recurrence of T-cells and macrophages across these tissues is therefore consistent with the established view that adaptive and innate immunity collaborate in chronic inflammation[11,14]. The presence of B-cells, dendritic cells, plasma cells, mast cells, and innate lymphoid cells further supports the complexity of the inflammatory environment.

This shared cellular pattern supports the comparative design of the thesis. The diseases differ in organ system, tissue architecture, and clinical phenotype, but the inflamed samples repeatedly contained immune populations associated with antigen presentation, cytokine production, and leukocyte recruitment[11,14]. If broadly similar immune-cell repertoires are present across diseases, then differences in network behaviour can be interpreted relative to signalling topology rather than as simple consequences of completely different cellular compositions. This does not mean that the diseases are immunologically identical; instead, it means that the same broad inflammatory cell classes recur in distinct tissue contexts.

The use of public scRNA-seq datasets was a practical strength as well as a limitation. A prospective study collecting matched primary tissue across twelve diseases would require many patient groups, distinct biopsy procedures, disease-specific control tissues, and harmonised sample handling across vascular, pancreatic, intestinal, lung, skin, salivary gland, synovial, and endometrial sites. Such an experiment would be difficult to deliver within a single project, especially where healthy tissue is ethically or practically hard to obtain. Harmonised reanalysis of public datasets therefore provides a feasible way to ask a cross-disease question, while retaining the limitation that

sampling design, sequencing platform, and control availability differed between source studies[25,39,42,46,50,151–154].

The fold-change and TP10K analyses showed that cytokine-axis activity varied by disease and by signalling pathway. IPF showed increased chemokine ligand and chemokine receptor expression, while Crohn's disease and ulcerative colitis showed increased NF-κB receptor expression; these patterns are consistent with published evidence that chemokines contribute to fibrotic lung inflammation and that NF-κB-linked cytokine signalling is prominent in intestinal inflammation[28,30,31,51]. More broadly, chemokine ligands and chemokine receptors, NF-κB pathway markers, and STAT pathway markers were detectable across the disease panel, supporting comparison of the three axes within the same inflammatory settings[57–60]. The relevant comparison is therefore not whether all diseases express the same cytokines at the same level, but whether the three signalling axes are represented in the diseased tissues being compared.

The ligand-receptor count analysis identified a consistent asymmetry in the chemokine axis. Inflamed tissues expressed substantially more chemokine ligands than receptors, whereas ligand and receptor diversity were more balanced for cytokines signalling through NF-κB and STAT. This transcriptional asymmetry is consistent with the known organisation of the chemokine system, in which many ligands bind overlapping sets of receptors and receptor-side promiscuity is common[58,61,78]. It is also consistent with evolutionary analyses showing ligand expansion with comparatively stable receptor diversity in the chemokine system[80]. The result does not itself prove redundancy, but it identifies the molecular arrangement that makes redundant connectivity plausible.

The ligand-receptor asymmetry also helps connect these expression results to the therapeutic problem. Chemokine-receptor promiscuity has been interpreted either as

functional redundancy or as context-dependent specialisation[57,87,88]. These transcriptional data leave that distinction unresolved by themselves, because binding potency, biased agonism, and spatial localisation require additional measurements. They do show, however, that human inflammatory tissues contain both ligand excess and the receptor repertoire needed to realise promiscuous ligand-receptor networks[61,78]. The next chapter therefore tests whether this molecular foundation produces measurable network redundancy.

The spatial single-cell RNA-seq result in this chapter has a specific role. GSE314158 psoriasis skin and GSE312415 ulcerative colitis rectal tissue were the two public spatial single-cell RNA-seq datasets that reported gene expression and tissue position for individual cells and measured enough chemokine, NF-κB, and STAT ligand and receptor genes for tile-level analysis[149,150]. Both datasets contained immune and non-immune cells, and the measured cytokine genes were sufficient for the tile-level network analysis in Chapter 4. Chapter 3 therefore establishes the spatial single-cell RNA-seq expression basis for the redundancy and robustness analysis in Chapter 4.

Two further limitations affect interpretation of the expression data. First, RNA abundance is a proxy for protein expression and functional activity; protein abundance can differ from transcript abundance because of protein stability, translational control, and post-translational regulation[57,141]. Second, the primary twelve-disease expression analysis used dissociated scRNA-seq datasets, whereas chemokine function depends on local concentration gradients and tissue architecture[57,91,92]. Cell-position information was available for the psoriasis skin and ulcerative colitis rectal tissue datasets, and CosMx 6K measured targeted gene panels in these two disease contexts[149,150].

These observations further support the mechanistic question raised above: why have therapies targeting the NF-κB and STAT axes achieved clinical success while chemokine-

targeted approaches have not[7–9]? Testing this hypothesis requires formal network analysis, which is the subject of the following chapter.

Chapter 4

Construction and Analysis of Chemokine Signalling Networks in Inflammatory Disease

4.1 Introduction

The failure of single-target chemokine receptor antagonists in inflammatory disease represents a systematic pattern rather than isolated clinical disappointments[7,8,45]. Over three decades, numerous antagonists targeting CCR1, CCR2, CCR3, CCR5, CCR9, CXCR2, and CXCR3 have entered clinical trials for rheumatoid arthritis, inflammatory bowel disease, and other immune-mediated conditions[8,45,100]. Nearly all have failed to demonstrate efficacy[7,8]. The consistency of these failures across different targets, diseases, and drug modalities suggests that the problem lies not with individual molecules but with the fundamental organisation of chemokine signalling in human disease.

Two competing hypotheses seek to explain this therapeutic failure[87,88]. The redundancy hypothesis proposes that the chemokine system is characterised by extensive ligand-receptor promiscuity that creates overlapping signalling pathways[57,86,87]. When one

pathway is blocked, alternative pathways maintain immune cell recruitment, rendering single-target inhibitors ineffective[87]. This hypothesis draws support from chemokine-receptor promiscuity: multiple chemokines bind more than one receptor, and multiple receptors bind more than one chemokine[61,78]. It also draws support from evolutionary considerations, since parasites such as ticks and viruses have independently evolved multi-chemokine inhibitors, suggesting that single-target blockade is insufficient to disable the system[6,94,95].

The alternative hypothesis proposes that chemokine-receptor interactions are context-specific rather than redundant[88,112,113]. According to this view, different ligand-receptor pairs produce distinct functional outcomes through biased agonism, differential expression kinetics, or spatial localisation[79,88]. Apparent redundancy in vitro or in knockout models reflects experimental simplifications that mask in vivo specificity[88,112]. Proponents cite evidence from mouse genetics: deletion of individual chemokines or receptors has produced distinct phenotypes across disease models, suggesting non-overlapping functions[88,93].

The debate has been difficult to resolve because most evidence derives from inbred mice maintained under pathogen-free conditions[114,115]. These animals lack the genetic diversity and immune experience of humans[114,115]. Inbred strains have near-identical inherited genetic backgrounds[114]. Pathogen-free housing results in naive-skewed immune systems with deficient effector and memory T-cell pools, whereas humans accumulate immune experience throughout life[115,116]. These differences may affect the degree of functional overlap between chemokine pathways: a redundant backup system may be unnecessary in a controlled laboratory environment but essential in the pathogen-rich world that shaped mammalian immune evolution[87].

Network biology provides a framework for quantifying redundancy and robustness in

complex systems[3,118]. Biological networks can be represented as graphs in which nodes are molecular components and edges are interactions[118]. Network-level properties such as redundancy and robustness emerge from the topology of these interactions and cannot be inferred from studying individual components in isolation[3,4]. Robustness, defined as the capacity to maintain function despite perturbation, is a property of evolved complex systems that arises through mechanisms including functional redundancy, modularity, and distributed control[3,105,106]. If the chemokine system is indeed redundant, this should be detectable as a network-level property.

This chapter describes the construction and analysis of cytokine-receptor interaction networks for the CK:CKR, NF-κB, and STAT signalling axes across twelve immune-mediated inflammatory diseases. Using graph-theoretic measures, I quantify edge redundancy, the degree to which networks contain more connections than minimally necessary for connectivity. Using simulated node removal, I assess functional and structural robustness, the capacity of networks to maintain receptor activation and structural integrity under progressive perturbation[124,125]. This systematic comparison across diseases and signalling axes tests whether the chemokine system possesses network properties that distinguish it from successfully targeted cytokine pathways.

Because chemokine function depends on cell position and tissue proximity, this chapter also analyses spatial scRNA-seq networks from psoriasis lesional skin and ulcerative colitis inflamed rectal tissue. This asks whether the higher CK:CKR redundancy observed in dissociated scRNA-seq data is also seen when ligand-receptor edges are restricted by spatial proximity.

4.2 Results

4.2.1 Construction of disease-specific cytokine networks

To identify the connections of cytokine ligands to immune cell receptors, ligand-receptor interaction networks were constructed. For each signalling category, CK:CKR, NF-κB, and STAT, expressed cytokine ligands and receptors were identified as those exceeding 0.005 transcript counts per 10 000, equivalent to 0.5 transcripts per million, in the disease scRNA-seq datasets. Ligands were identified from all tissue cells because cytokines can be produced by immune and non-immune cells. Receptors were identified only from immune cells because the analysis focuses on signalling into immune populations that mediate inflammatory recruitment and activation.

Known ligand-receptor interactions from the IUPHAR/BPS Guide to PHARMACOLOGY database were then incorporated to build 36 bipartite ligand-receptor networks, comprising 12 diseases multiplied by 3 signalling mechanisms[76]. These bipartite networks were ensured to have no isolated ligands or receptors. The same network-construction rules were applied to all three signalling axes.

Across the constructed networks, CK:CKR networks contained a median of 26.5 ligand-receptor edges, compared with medians of 16.0 for NF-κB and 7.0 for STAT. CK:CKR networks also had a higher median ligand count than the other axes, consistent with the ligand-heavy expression pattern described in Chapter 3. These summary values show that the chemokine networks had both a larger ligand repertoire and more curated ligand-receptor connections.

The chord diagram layout is summarised in **Figure 4.1**. It illustrates the three signalling axes in representative disease networks. Networks for all 12 diseases are shown in Appendix A.

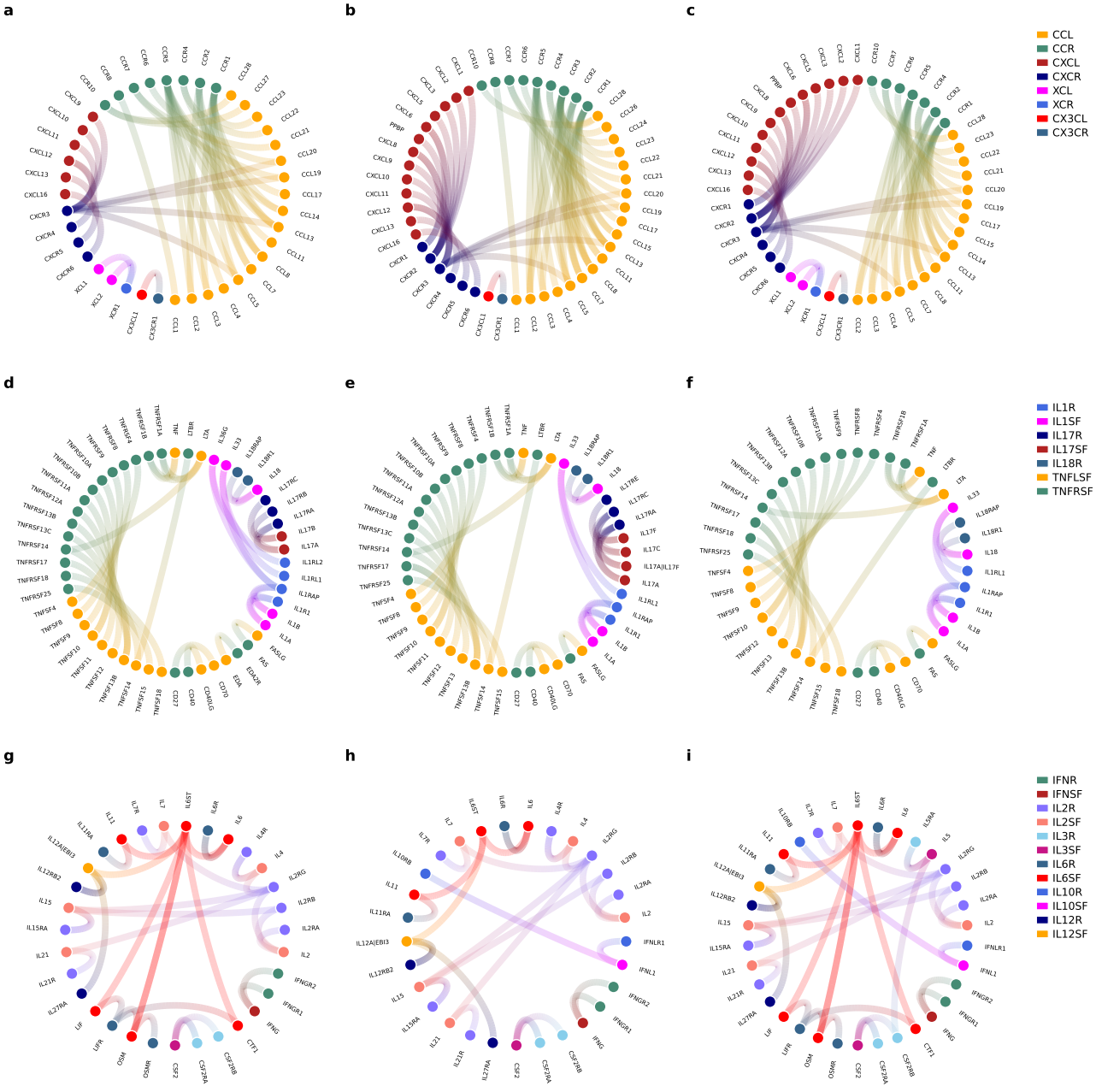

Figure 4.1: Chord diagram visualisation of disease-specific ligand-receptor networks. Panels **a**–**c** show CK:CKR networks, panels **d**–**f** show NF-κB networks, and panels **g**–**i** show STAT networks. Panels **a**, **d**, and **g** show systemic lupus erythematosus skin tissue, panels **b**, **e**, and **h** show Crohn's disease colon tissue, and panels **c**, **f**, and **i** show idiopathic pulmonary fibrosis lung tissue. Nodes arranged around each circumference represent ligands and receptors that passed disease-state expression filtering, and chords connect ligand-receptor pairs with curated activating interactions. Node and chord colours denote ligand or receptor families as indicated by the panel keys.

In each chord, ligands are represented on one side, receptors on the other, and each edge corresponds to a curated activating relationship that is possible because both partners are expressed in the disease tissue. A receptor connected to several ligands illustrates receptor-side promiscuity, while a ligand connected to several receptors illustrates ligand-side promiscuity.

4.2.2 Quantification of network edge redundancy

To quantify edge redundancy, the observed number of ligand-receptor edges and the minimum number required to cover every node were computed for each network. The minimal network for each condition was calculated using the minimum edge cover, which identifies the smallest set of edges that includes every node at least once.

This definition was chosen because it directly addresses whether a network contains more edges than are required to keep all expressed ligands and receptors represented. In a bipartite ligand-receptor graph, the minimum edge cover preserves at least one connection for every molecule in the network. Edges beyond this minimum represent alternative connections that are not needed for simple coverage of the node set. These additional edges provide alternative connections if an individual ligand-receptor interaction is lost.

E was defined as the observed ligand-receptor interaction set and $E_{\min}$ as a minimum edge cover. The latter is the smallest subset needed to include every node. Their sizes are $|E|$ and $|E_{\min}|$, respectively. To allow comparisons between networks of differing sizes, the normalised edge redundancy index R was defined as:

$$R = 1 - \frac{|E_{\min}|}{|E|} \tag{4.1}$$

This index ranges from 0, indicating no excess edges beyond those required to ensure

every node has at least one connection, to values approaching 1, indicating many edges beyond the minimum edge cover. Higher values therefore represent greater topological redundancy.

Across all 36 networks, the redundancy index was higher for CK:CKR networks than for NF-κB or STAT networks (**Figure 4.2**). This result held across all twelve diseases, indicating that elevated chemokine network redundancy is a general feature of immune-mediated inflammatory disease rather than a property of specific conditions.

The median redundancy index was 0.371 for CK:CKR networks, compared with 0.124 for NF-κB networks and 0.053 for STAT networks. CK:CKR redundancy exceeded NF-κB redundancy in all twelve diseases and exceeded STAT redundancy in all twelve diseases. The CK:CKR values ranged from 0.298 to 0.426, whereas several NF-κB and STAT networks had redundancy indices of zero, meaning that their observed edge sets were already close to the minimum needed to cover all nodes. Thus, the difference between axes was not driven by a single outlying disease; it was a repeated pattern across the disease panel.

The networks with zero redundancy are especially informative because they illustrate the opposite end of the redundancy-index range. In those graphs, each observed edge is required to keep every expressed node represented in the minimum edge cover. Removing an edge from such a network causes at least one ligand or receptor to lose coverage. By contrast, a CK:CKR network with a redundancy index near 0.4 contains a substantial fraction of edges that exceed this minimum coverage requirement. This supports a graph-topology claim rather than biological equivalence among those edges: the chemokine graph has more alternative connections available under the same topological rule.

The result also shows why ligand count alone is not enough. CK:CKR networks

had more ligands than receptors, but redundancy depends on how those ligands are distributed across receptors. If every chemokine ligand connected to a unique receptor, the network would remain sparse despite ligand excess. The observed redundancy indicates that expressed chemokines converge onto shared receptors and that some receptors retain multiple possible ligand inputs.

4.2.3 Promiscuity drives network redundancy

To test whether promiscuity explained variation in network redundancy, mean ligand promiscuity, mean receptor promiscuity, and overall promiscuity per network were quantified. Promiscuity was defined as the number of partners connected to each ligand or receptor.

Network edge redundancy strongly correlated with overall promiscuity, with a correlation coefficient of approximately 0.97. Receptor promiscuity showed a stronger relationship with redundancy, with a correlation of approximately 0.95, than ligand promiscuity, which showed a correlation of approximately 0.75. This correlation was driven by the high levels of receptor and overall promiscuity within the CK:CKR networks.

These results indicate that the number of partners per node, and especially receptor-side degree, largely explains variation in global network edge redundancy (**Figure 4.3**). Chemokine receptors bind many ligands, and this receptor-side promiscuity creates excess edges beyond the minimum required to cover all expressed nodes.

The distinction between ligand-side and receptor-side promiscuity matters for interpreting the redundancy index. Ligand-side promiscuity means that one cytokine is connected to several receptors, whereas receptor-side promiscuity means that one receptor is connected to several ligands. The stronger relationship between receptor

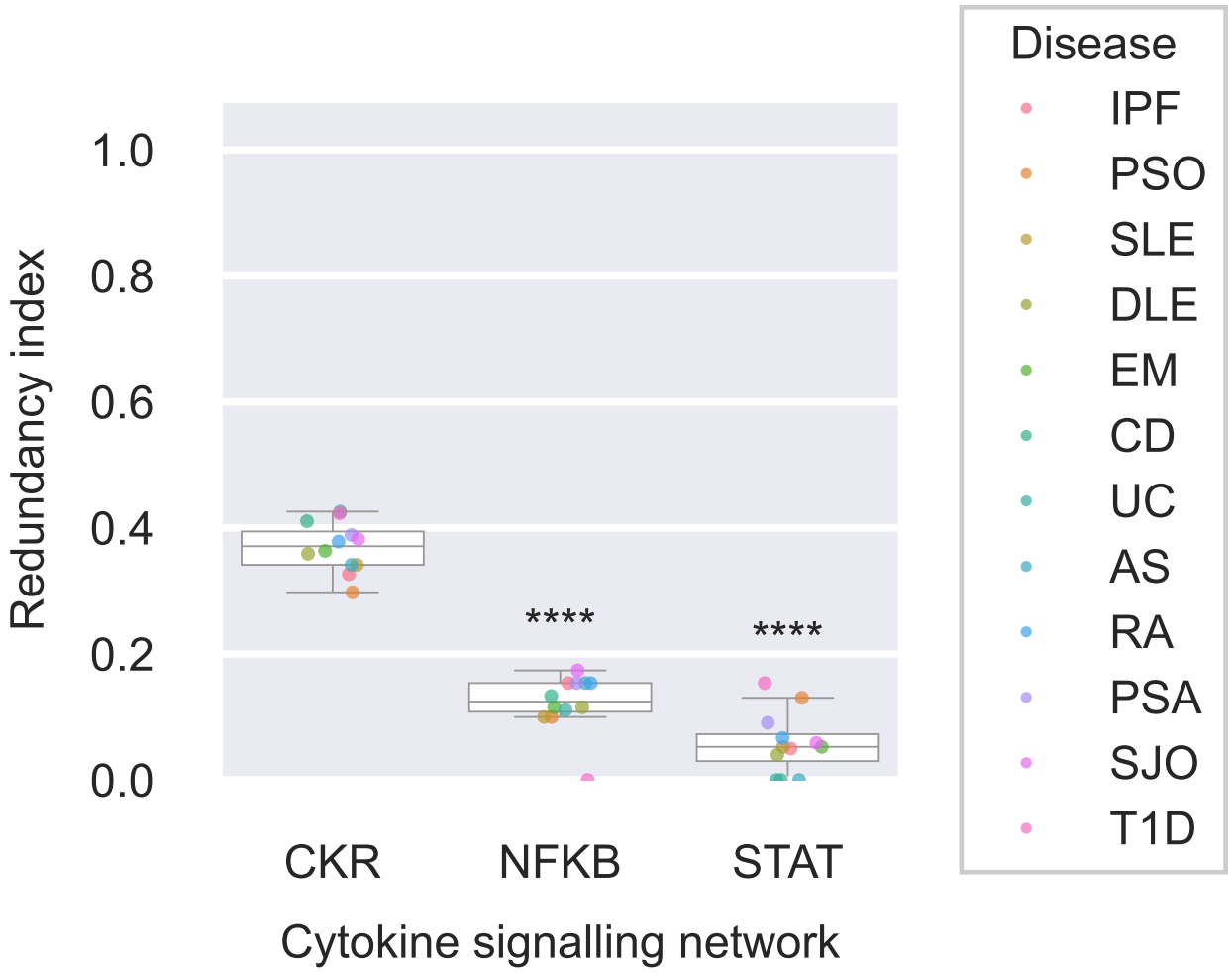

Figure 4.2: Edge redundancy across the 36 disease-specific cytokine signalling networks. The redundancy index was defined as $R = 1 - |E_{\min}|/|E|$, where $|E|$ is the observed number of ligand-receptor edges and $|E_{\min}|$ is the size of the minimum edge cover. Networks are grouped by signalling axis: CK:CKR, NF-κB, and STAT. Each coloured point represents one disease-specific network, with colours denoting diseases as shown in the legend. Boxes show the interquartile range and median across diseases for each signalling axis. Asterisks indicate Holm-adjusted Mann-Whitney U tests comparing CK:CKR networks with NF-κB or STAT networks, with **** indicating $p < 0.0001$.

promiscuity and redundancy shows that receptor-side multi-partner connectivity is the main topological contributor to the CK:CKR pattern.

The correlation with overall promiscuity also provides a useful internal check on the redundancy metric. Redundancy should increase when nodes have more partners, because additional partners create additional edges beyond the minimum needed for node coverage. The observed relationship confirms that the redundancy index is capturing the intended topological feature rather than an unrelated property of network size. At the same time, the analysis separates size from redundancy: a larger network is not automatically more redundant unless its nodes have excess alternative connections.

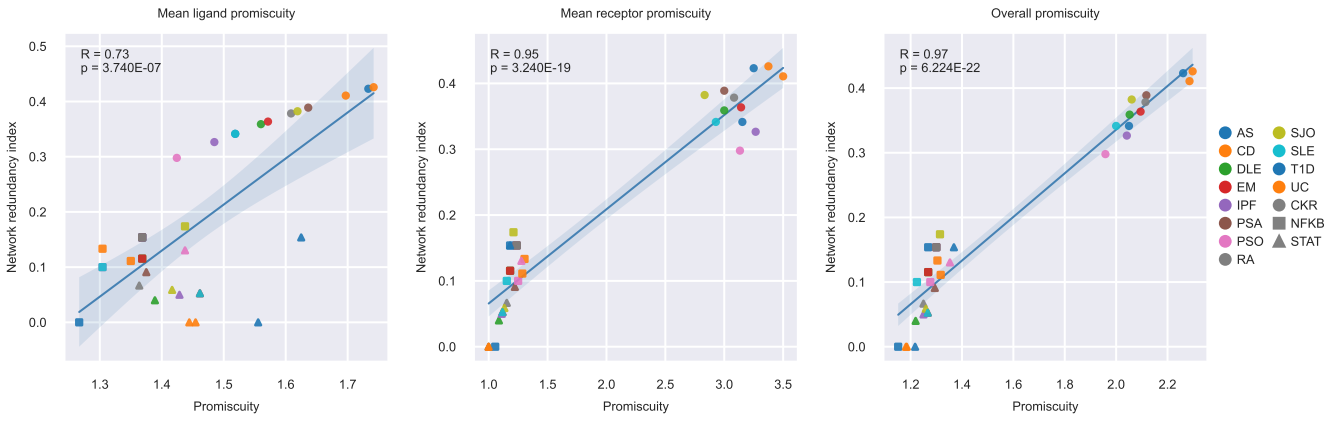

Figure 4.3: Relationship between edge redundancy and node promiscuity across the 36 disease-specific cytokine signalling networks. The y-axis shows the network redundancy index. The x-axis shows mean ligand promiscuity, mean receptor promiscuity, or overall promiscuity, where promiscuity is the mean number of ligand-receptor partners per node in the indicated partition or across the whole bipartite network. Each point represents one disease-specific network. Point colour denotes disease and point shape denotes signalling axis: CK:CKR, NF-κB, or STAT. Blue lines show linear regression fits with shaded 95% confidence intervals. Inset values show Pearson correlation coefficients and associated two-sided p values for each promiscuity measure.

4.2.4 Functional robustness of disease networks

The robustness of each disease network was next assessed by simulating random ligand loss. Robustness was evaluated by sequentially removing cytokine ligand nodes and quantifying the effect on network properties[125]. In this approach, cytokine ligands were removed in random order, while receptors were considered lost once they lost all connections to remaining ligands. An example of this process for the SLE NF-κB network is shown in **Figure 4.4**.

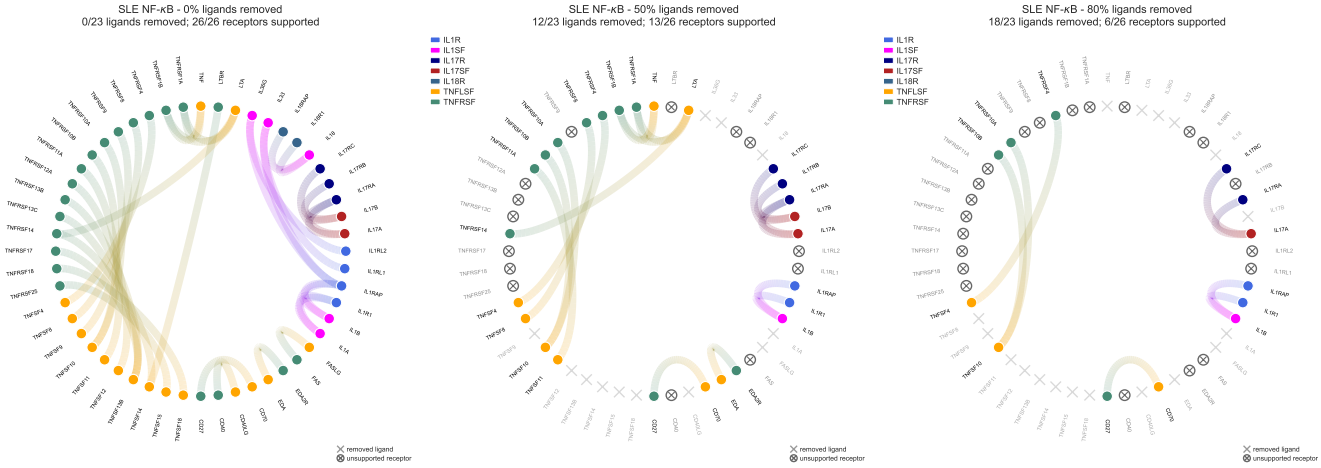

Figure 4.4: Example of sequential cytokine-ligand removal in the systemic lupus erythematosus NF-κB-axis network. Panels show the same network after 0%, 50%, and 80% of ligand nodes were removed using a fixed random removal order. Nodes arranged around each circumference represent ligands and receptors, and ribbons show remaining curated ligand-receptor interactions. Grey crosses indicate removed ligands. Circled crosses indicate unsupported receptors, defined as receptors with no remaining ligand connections after ligand removal. Panel titles report the number of ligands removed and the number of receptors that remain supported by at least one ligand. Node and ribbon colours denote ligand or receptor families as indicated by the keys.

For each network, 250 random cytokine-removal sequences were performed and robustness was estimated using secondary extinction analysis, which is widely applied in ecological network studies[124]. Secondary extinction robustness was defined as the

area under the extinction curve plotting the percentage of persisting receptors against the percentage of ligands removed. This metric quantifies functional robustness at the receptor level, assuming that a receptor remains functional as long as it retains at least one activating ligand.

In graph terms, primary extinction corresponds to removal of a cytokine ligand from the network. Secondary extinction occurs when a receptor loses its final remaining ligand and therefore no longer has an activating cytokine input under the graph model. A network in which receptors have several ligand alternatives will show delayed secondary extinction, because receptors persist until all their ligands have been removed. This is the expected behaviour of a receptor-promiscuous network.

Example median extinction curves for Crohn's disease and SLE showed that in each case the CK:CKR curve enclosed the largest area (**Figure 4.5**). Larger enclosed areas indicate greater receptor persistence during ligand removal.

Across all 36 networks, CK:CKR networks exhibited significantly higher secondary-extinction robustness than NF-κB or STAT networks (**Figure 4.6**). Higher values indicate delayed receptor loss during ligand removal.

The median secondary-extinction robustness was 0.660 for CK:CKR networks, compared with 0.535 for NF-κB and 0.517 for STAT. CK:CKR robustness exceeded both comparator axes in all twelve diseases. This means that, under random ligand removal, chemokine receptors retained at least one remaining ligand for longer than receptors in the other two signalling systems. The result is the functional counterpart of the redundancy analysis: the excess edges in CK:CKR networks translate into persistence of receptor connectivity under simulated perturbation.

The shape of the extinction curve is also interpretable. A steep early decline would indicate that receptors are dependent on a small number of ligands and are rapidly

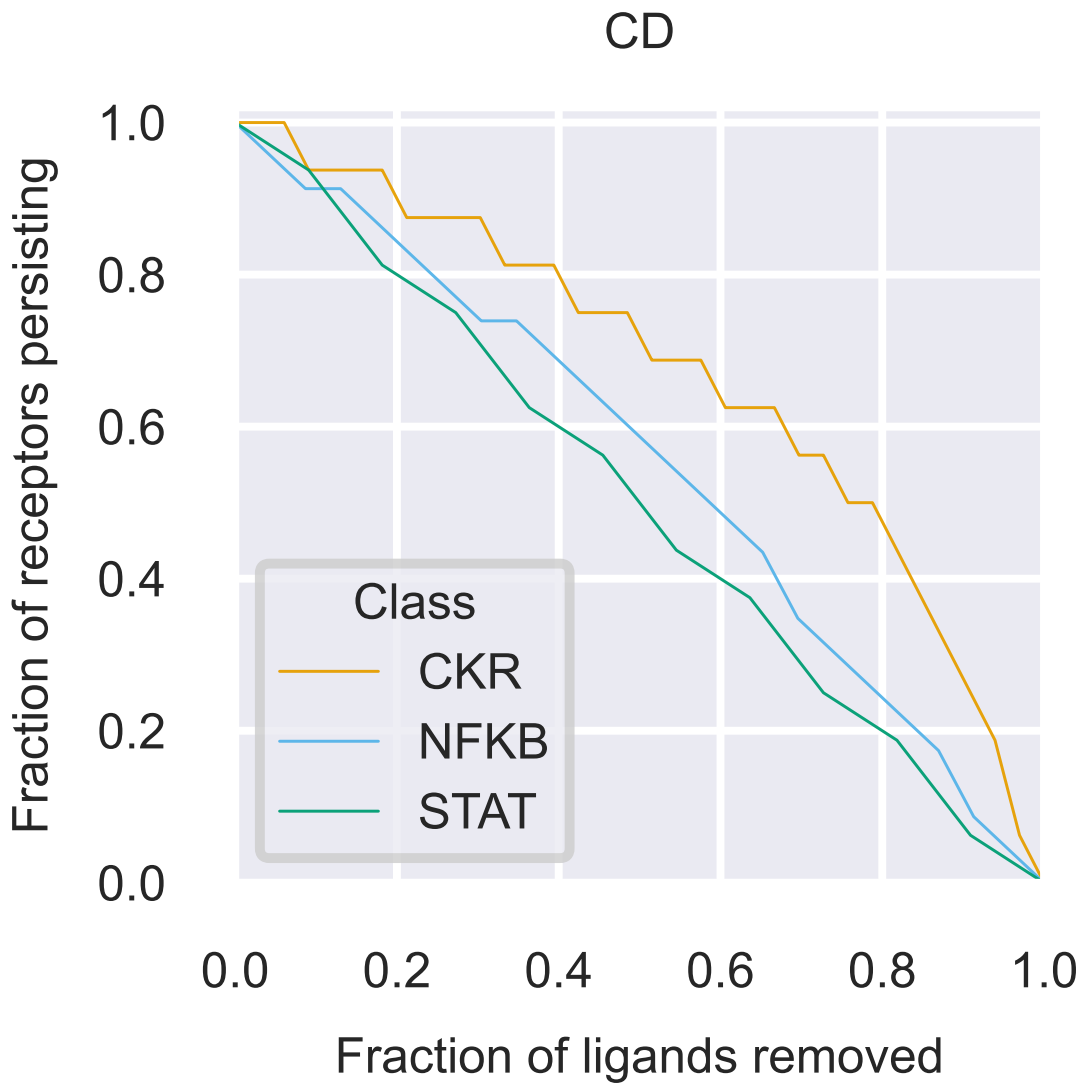

Figure 4.5: Secondary extinction robustness curves for Crohn's disease cytokine signalling networks. Ligands were removed sequentially in 250 random removal orders for each signalling axis. The x-axis shows the fraction of ligand nodes removed, and the y-axis shows the fraction of receptor nodes that remained supported by at least one ligand. Lines show the median receptor persistence curve for CK:CKR, NF-κB, and STAT networks across removal simulations. The area under each curve was used to quantify functional robustness, with larger values indicating delayed receptor loss during ligand removal.

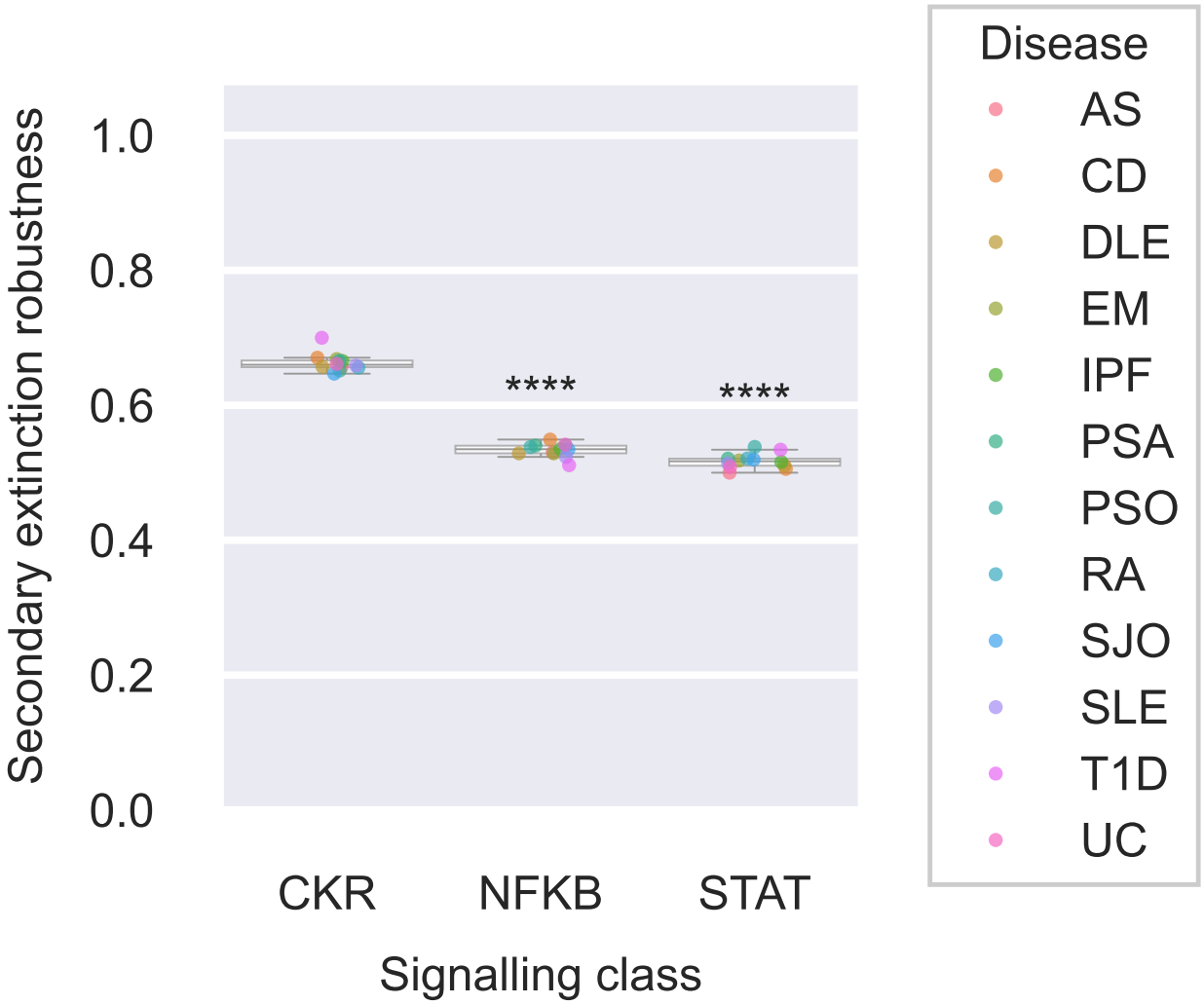

Figure 4.6: Secondary extinction robustness across the 36 disease-specific cytokine signalling networks. Robustness was quantified as the area under the secondary extinction curve, where each curve plots the fraction of receptor nodes remaining supported against the fraction of ligand nodes removed during random sequential ligand removal. Networks are grouped by signalling axis: CK:CKR, NF-κB, and STAT. Each coloured point represents one disease-specific network, with colours denoting diseases as shown in the legend. Boxes show the interquartile range and median across diseases for each signalling axis. Asterisks indicate Holm-adjusted Mann-Whitney U tests comparing CK:CKR networks with NF-κB or STAT networks, with **** indicating $p < 0.0001$. Higher values indicate greater functional robustness to ligand loss.

lost when those ligands are removed. A flatter curve indicates that receptors persist after many ligands have been removed, which is expected when receptors have multiple ligand alternatives.

4.2.5 Structural robustness of disease networks

Largest connected component robustness was also computed for each disease network[125]. After each ligand-removal step, the number of original graph nodes present in the largest connected component was recorded as a fraction of the initial node count. The area under this curve summarises how much of the original graph remains in a single connected component during progressive ligand removal.

This second analysis records a different output from the same ligand-removal simulation. Secondary extinction counts receptors that retain at least one ligand. Largest connected component robustness records the fraction of the original node set contained in the largest connected component after each ligand-removal step. It therefore summarises graph fragmentation without treating receptor persistence as the endpoint.

Example median largest connected component robustness curves for Crohn's disease and SLE showed that after each ligand-removal step, a larger fraction of original CK:CKR graph nodes remained in the largest connected component than in the NF-κB or STAT networks (**Figure 4.7**). This pattern indicates slower fragmentation of the CK:CKR networks.

Across all 36 networks, CK:CKR networks exhibited significantly higher largest connected component robustness than NF-κB or STAT networks (**Figure 4.8**). The robustness results were consistent across the different expression thresholds used. Taken together, these analyses indicate that CK:CKR signalling networks in human disease exhibit significantly greater secondary-extinction and largest-connected-component

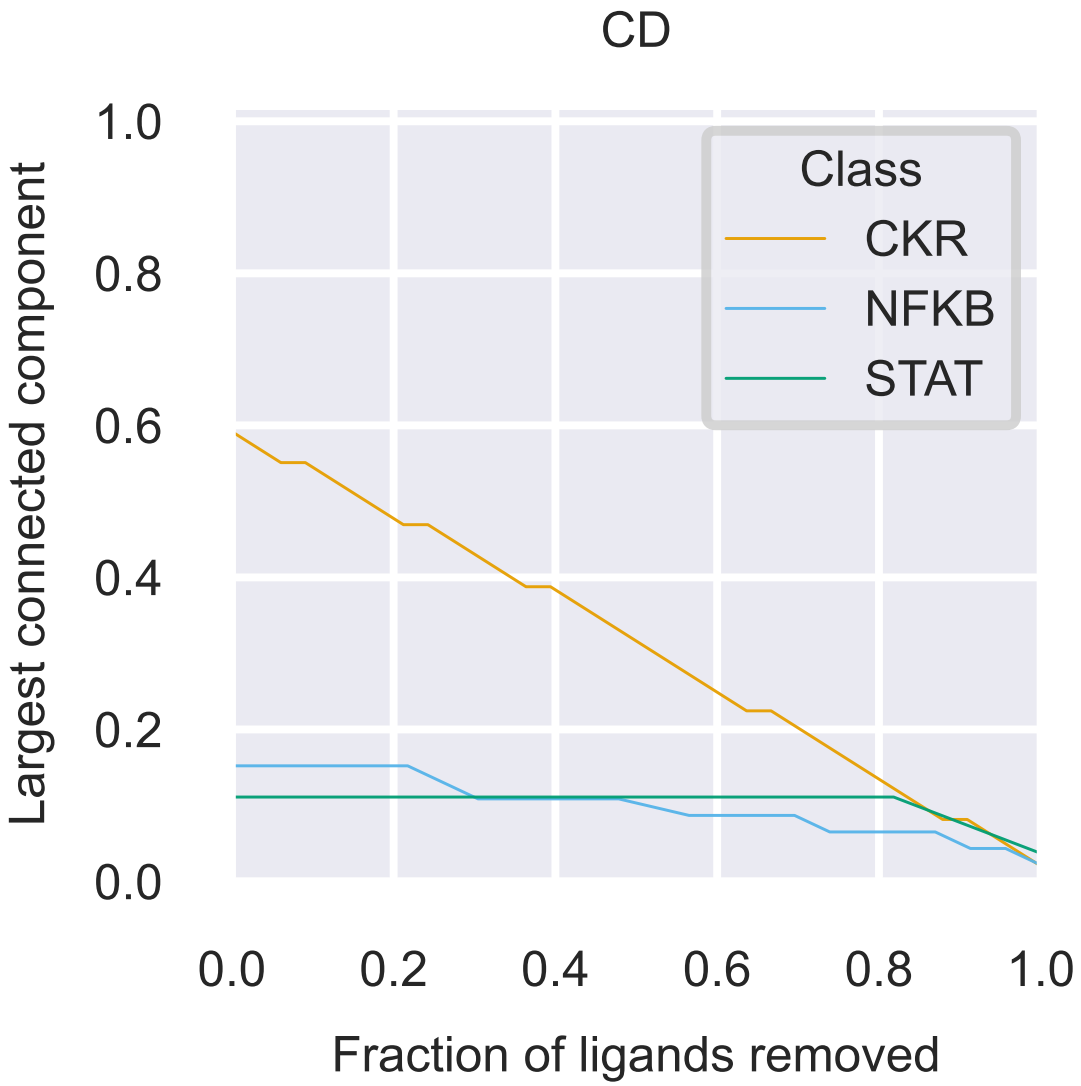

Figure 4.7: Largest connected component robustness curves for Crohn's disease cytokine signalling networks. Ligands were removed sequentially in 250 random removal orders for each signalling axis. The x-axis shows the fraction of ligand nodes removed, and the y-axis shows the fraction of original network nodes retained in the largest connected component after each removal step. Lines show the median largest-connected-component curve for CK:CKR, NF-κB, and STAT networks across removal simulations. The area under each curve was used to quantify structural robustness, with larger values indicating slower fragmentation of the network during ligand removal.

robustness than NF-κB or STAT networks.

The median largest connected component robustness value was 0.309 for CK:CKR networks, compared with 0.104 for NF-κB and 0.116 for STAT. As with secondary extinction, CK:CKR networks had higher values than both comparator axes in all twelve diseases. The agreement between the two robustness analyses strengthens the conclusion because the metrics summarise different graph properties. Secondary extinction is calculated from receptors retaining at least one ligand. Largest connected component robustness is calculated from the largest connected component size after each ligand-removal step, normalised to the initial node count. Both identify the chemokine axis as the most robust signalling network in this comparison.

The largest connected component result is not redundant with the secondary-extinction result. A receptor can retain one ligand while the remaining graph is split into disconnected components. Conversely, a large connected component can persist while individual receptors lose all ligand connections. Observing higher CK:CKR values in both analyses therefore indicates that chemokine networks preserve both receptor-level connectivity and a larger connected component under random ligand loss.

4.2.6 Sensitivity analysis of expression thresholds

The analysis was repeated with expression cutoffs of 0.01 and 0.05 transcripts per 10 000 to test whether gene inclusion affected the measured redundancy. The edge redundancy results were consistent across the different expression thresholds used. This robustness to threshold choice indicates that the observed differences between CK:CKR and NF-κB/STAT networks are not artefacts of the expression cutoff.

The sensitivity analysis addressed the possibility that a permissive expression threshold could inflate network connectivity by retaining very low-abundance transcripts.

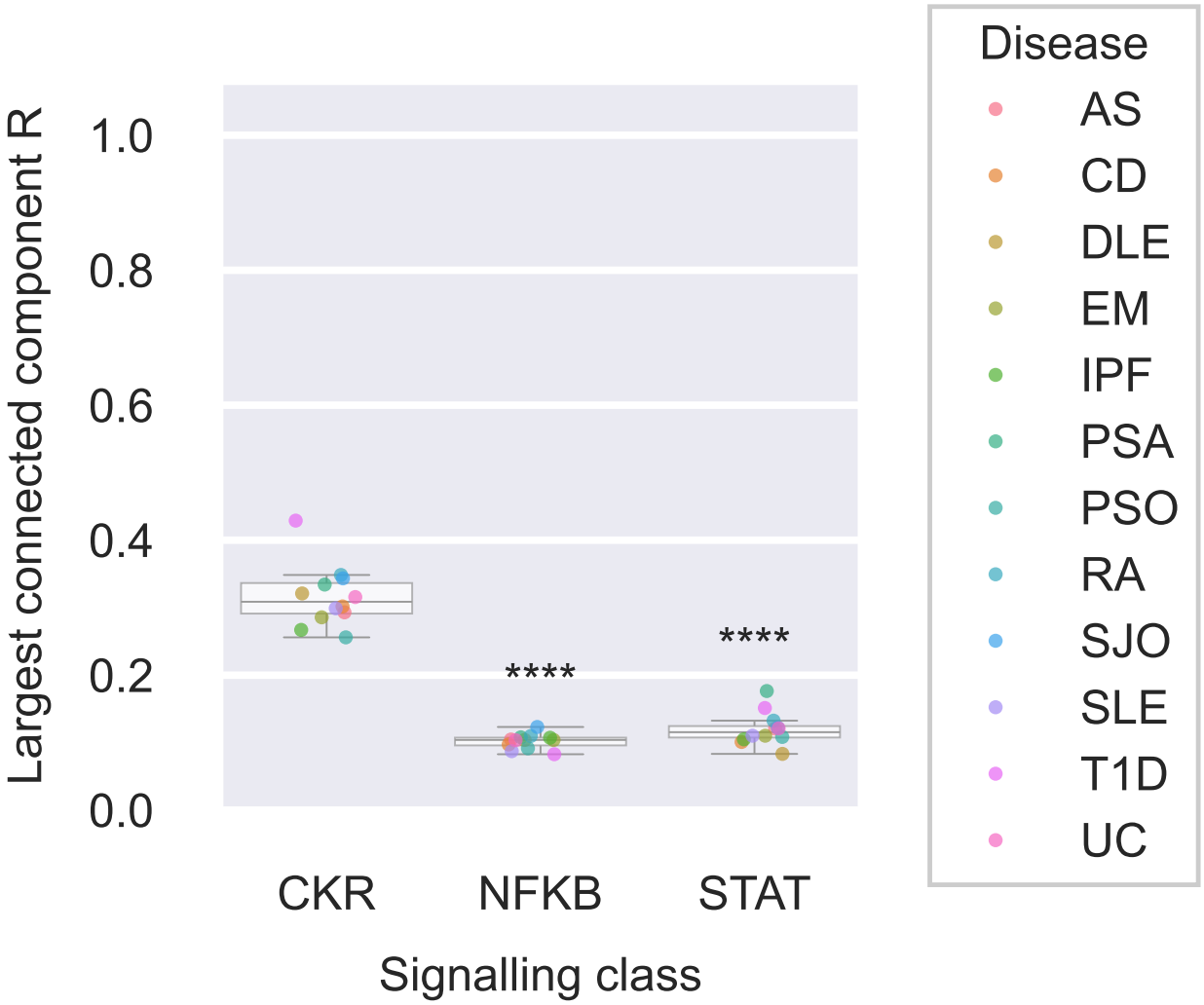

Figure 4.8: Largest connected component robustness across the 36 disease-specific cytokine signalling networks. Robustness was quantified as the area under the largest connected component curve, where each curve plots the fraction of original network nodes retained in the largest connected component against the fraction of ligand nodes removed during random sequential ligand removal. Networks are grouped by signalling axis: CK:CKR, NF-κB, and STAT. Each coloured point represents one disease-specific network, with colours denoting diseases as shown in the legend. Boxes show the interquartile range and median across diseases for each signalling axis. Asterisks indicate Holm-adjusted Mann-Whitney U tests comparing CK:CKR networks with NF-κB or STAT networks, with **** indicating $p < 0.0001$. Higher values indicate greater structural robustness to ligand loss.

Raising the threshold removes weakly expressed ligands and receptors and therefore tests whether the CK:CKR result depends on marginal expression. The persistence of the same qualitative pattern at higher thresholds shows that elevated CK:CKR redundancy is not simply a consequence of including more borderline-expressed chemokines (**Figure 4.9**).

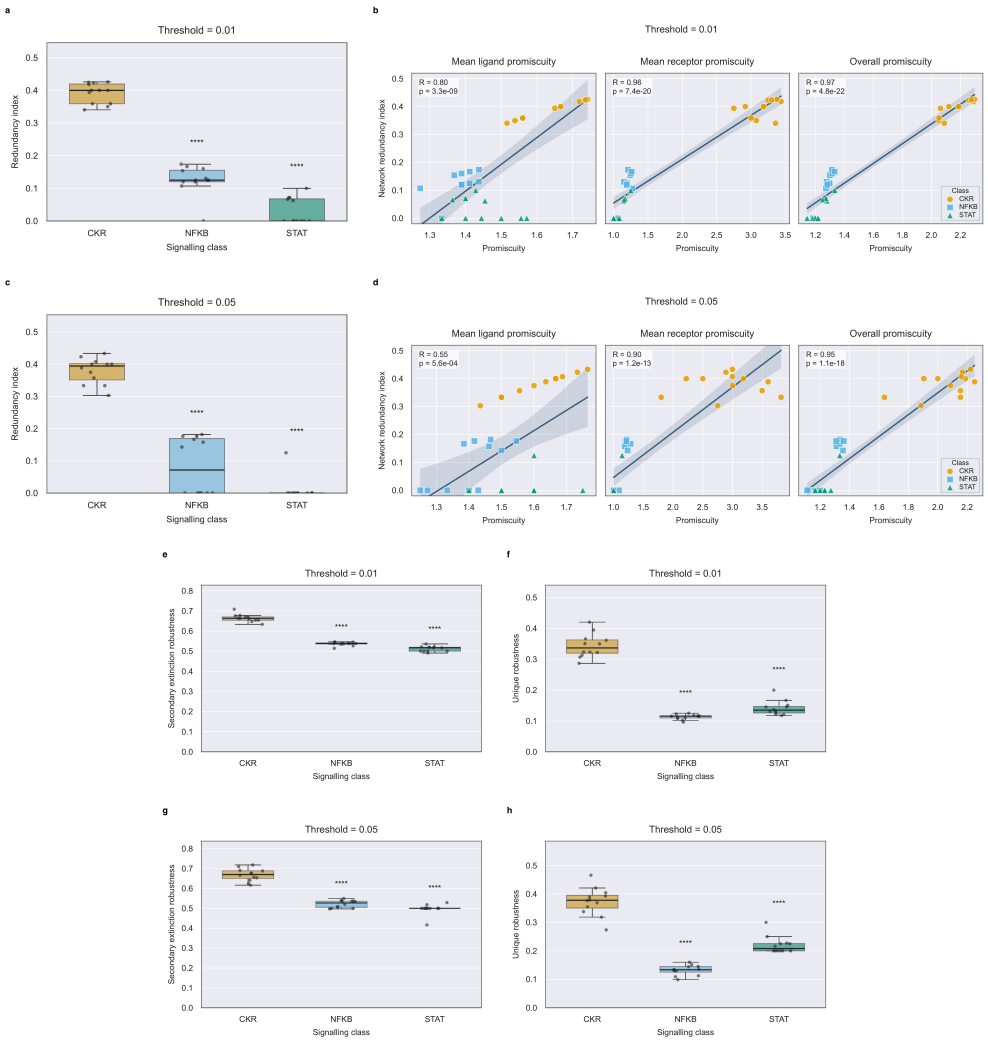

Figure 4.9: Sensitivity analysis showing the effect of expression-filtering threshold on network metrics. Networks were reconstructed using disease-state expression thresholds of 0.01 and 0.05 transcripts per 10,000. Panels **a** and **c** show edge redundancy index at thresholds 0.01 and 0.05, respectively. Panels **b** and **d** show the relationship between redundancy and mean ligand promiscuity, mean receptor promiscuity, or overall promiscuity at the same thresholds. Panels **e** and **g** show secondary-extinction robustness, and panels **f** and **h** show largest connected component robustness, at thresholds 0.01 and 0.05, respectively. Each point represents one disease-specific network. Colours denote diseases, and point shape denotes signalling axis in the promiscuity panels. Boxplots show medians and interquartile ranges. Blue lines show linear regression fits with shaded 95% confidence intervals, and inset values show Pearson correlation coefficients with associated p values. Asterisks denote pairwise Mann-Whitney U comparisons of CK:CKR networks with NF-κB or STAT networks, with **** indicating $p < 0.0001$.

4.2.7 Spatial single-cell RNA-seq network analysis in psoriasis and ulcerative colitis

Two public spatial scRNA-seq datasets were used for network analysis: psoriasis skin and ulcerative colitis rectal tissue[149,150]. Both datasets recorded expression and coordinates for individual cells and covered sufficient ligand and receptor genes from the chemokine, NF-κB, and STAT axes to support tile-level networks. In this analysis, a tile-level ligand-receptor interaction was defined as one curated ligand-receptor pair for which at least one cell in a square tile expressed the ligand and at least one immune cell in the same tile expressed the receptor. The primary setting used 500 μm square tiles and the 0.005 TP10K expression threshold.

After curated ligand-receptor interactions were restricted to genes measured by the spatial scRNA-seq panel, 62 CK:CKR, 30 NF-κB, and 40 STAT interactions could be evaluated in both disease tissues. At 1500 μm square tile size, tile-level ligand-receptor networks were detected for each signalling axis in psoriasis lesional skin and ulcerative colitis inflamed rectal tissue (**Figure 4.10**). This coarse tile size was used to show the tissue coordinate plane, square-tile structure, and ligand-receptor network structure together.

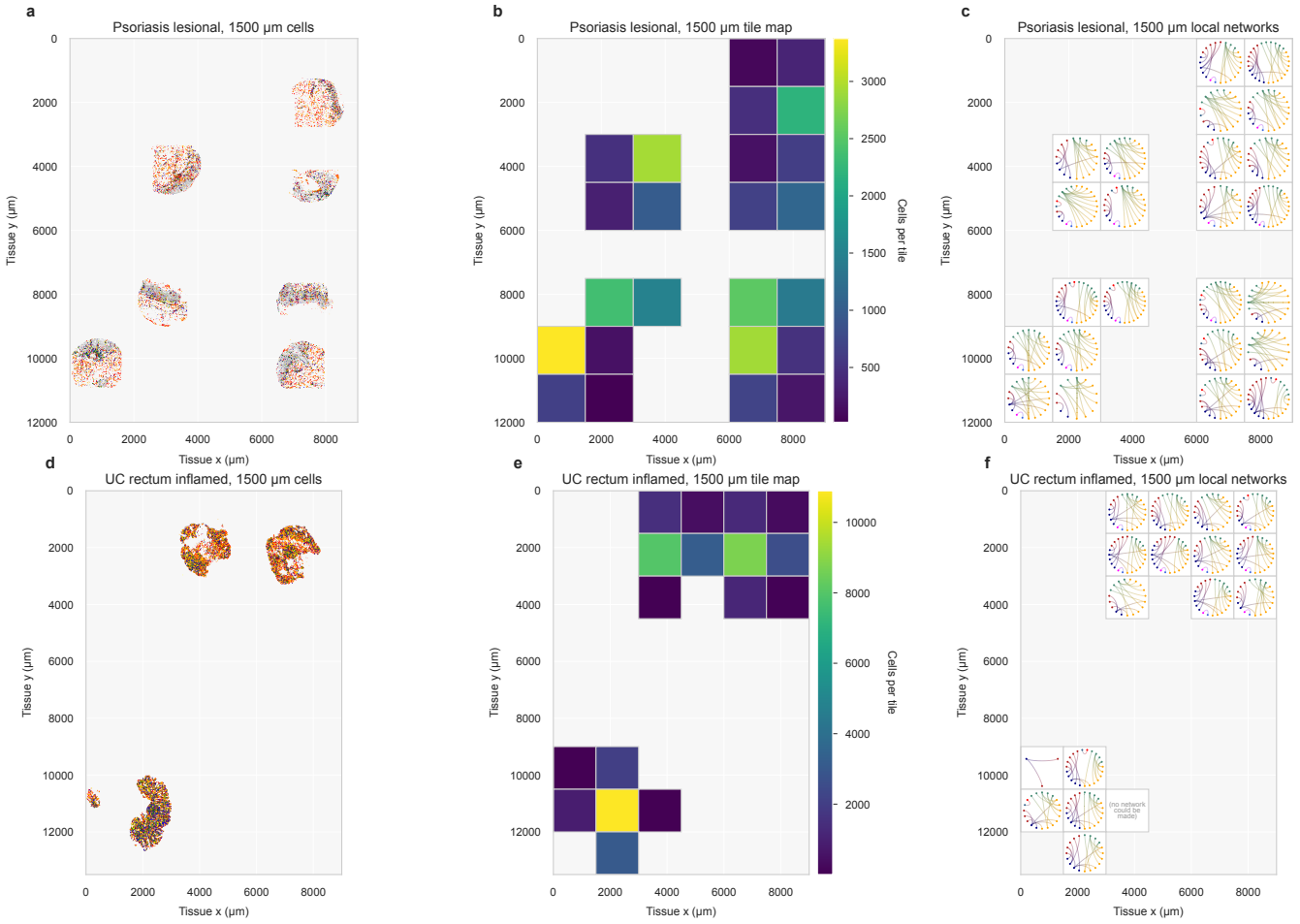

Figure 4.10: Tile-level ligand-receptor networks for psoriasis lesional skin and ulcerative colitis inflamed rectal tissue at 1500 μm square tile side length. Panels **a**, **b**, and **c** show psoriasis lesional skin cell positions, cell counts in square tiles, and ligand-receptor networks placed inside the square tiles where they were detected. Panels **d**, **e**, and **f** show ulcerative colitis inflamed rectal tissue cell positions, cell counts in square tiles, and ligand-receptor networks placed inside the square tiles where they were detected. In panels **c** and **f**, nodes represent measured ligand and receptor genes that passed the expression threshold, and edges represent curated ligand-receptor pairs detected within a square tile. Panels **c** and **f** show CK:CKR interactions. Ligand nodes represent CCL, CXCL, CX3CL1 and XCL1/2 chemokines; receptor nodes represent CCR, CXCR, CX3CR1 and XCR1 chemokine receptors.

Tile-level networks were constructed for CK:CKR, NF-κB, and STAT in both disease tissues across the tested tile sizes. An active tile was a square tile containing at least one tile-level ligand-receptor interaction for the specified signalling axis. Here, one

tile-level ligand-receptor interaction was defined as one ligand-receptor pair detected in one tile, with at least one cell expressing the ligand and at least one immune cell expressing the receptor. At the primary 500 μm tile size and 0.005 TP10K threshold, psoriasis lesional skin produced 8 276 tile-level ligand-receptor interactions. These comprised 4 494 CK:CKR interactions across 82 active tiles, 1 823 NF-κB interactions across 83 active tiles, and 1 959 STAT interactions across 81 active tiles. Ulcerative colitis inflamed rectal tissue produced 6 469 tile-level ligand-receptor interactions, comprising 3 357 CK:CKR interactions across 58 active tiles, 1 438 NF-κB interactions across 57 active tiles, and 1 674 STAT interactions across 57 active tiles.

Across active tiles at the primary 500 μm tile size, CK:CKR had higher median redundancy and robustness values than NF-κB or STAT in both disease tissues. In psoriasis lesional skin, median CK:CKR redundancy was 0.4194, median secondary extinction AUC was 0.6765, and median largest connected component AUC was 0.3189. The corresponding NF-κB values were 0.1481, 0.5514, and 0.1291, and the corresponding STAT values were 0.08333, 0.5404, and 0.1207. In ulcerative colitis inflamed rectal tissue, median CK:CKR redundancy was 0.4194, median secondary extinction AUC was 0.6773, and median largest connected component AUC was 0.3181. The corresponding NF-κB values were 0.1429, 0.5534, and 0.1125, and the corresponding STAT values were 0.09375, 0.5410, and 0.1145.

The distribution of tile-level metrics showed higher CK:CKR values than NF-κB and STAT in both spatial scRNA-seq datasets (**Figure 4.11**). This ordering was observed for redundancy, secondary extinction AUC, and largest connected component AUC. CK:CKR was higher than both comparator axes in psoriasis lesional skin and ulcerative colitis inflamed rectal tissue.

Square side lengths from 100 to 1500 μm were evaluated every 50 μm. Expression

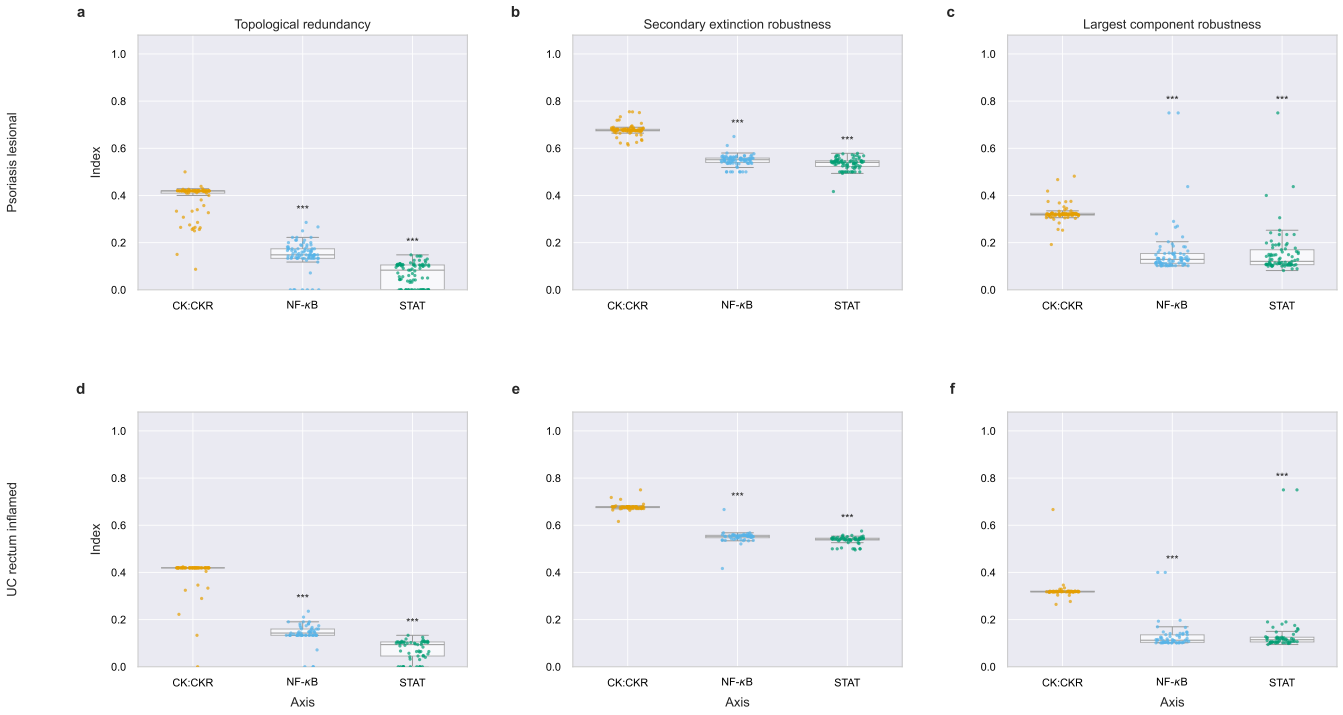

Figure 4.11: Tile-level redundancy and robustness distributions for CK:CKR, NF-κB, and STAT networks in psoriasis lesional skin and ulcerative colitis inflamed rectal tissue. Panels **a**, **b**, and **c** show psoriasis lesional skin redundancy index, secondary extinction AUC, and largest connected component AUC. Panels **d**, **e**, and **f** show ulcerative colitis inflamed rectal tissue redundancy index, secondary extinction AUC, and largest connected component AUC. Values were calculated from tile-level ligand-receptor networks constructed from cells in the same square tile.

thresholds of 0.005, 0.01, and 0.05 TP10K were tested separately (**Figure 4.12**). Median redundancy, secondary extinction AUC and largest connected component AUC were identical across the three thresholds at each tile size in both disease tissues. At the 0.005 TP10K threshold, the fraction of square tiles containing at least one tile-level ligand-receptor interaction increased as tile size increased. In psoriasis lesional skin, these fractions at 500 μm were 0.911 for CK:CKR, 0.922 for NF-κB, and 0.900 for STAT. In ulcerative colitis inflamed rectal tissue, the corresponding fractions were 0.951, 0.934, and 0.934. Across the full tile-size and expression-threshold sensitivity analysis, CK:CKR values remained higher than NF-κB and STAT values

for redundancy and both robustness metrics.

Pairwise Wilcoxon signed-rank tests with Benjamini-Hochberg correction supported higher CK:CKR values than NF-κB and STAT values at selected tile sizes across all three tested expression thresholds (**Figure 4.13**). For CK:CKR compared with NF-κB or STAT, all adjusted p values at the primary 500 μm tile size and 0.005 TP10K threshold were below 6.4×10^{-13} in psoriasis lesional skin and below 4.5×10^{-8} in ulcerative colitis inflamed rectal tissue. Paired differences and adjusted p values were identical across the three thresholds at each tile size in both tissues. The spatial scRNA-seq network analysis showed the same ordering observed in the primary twelve-disease scRNA-seq network analysis: CK:CKR networks had higher redundancy, secondary extinction robustness, and largest connected component robustness than NF-κB and STAT networks.

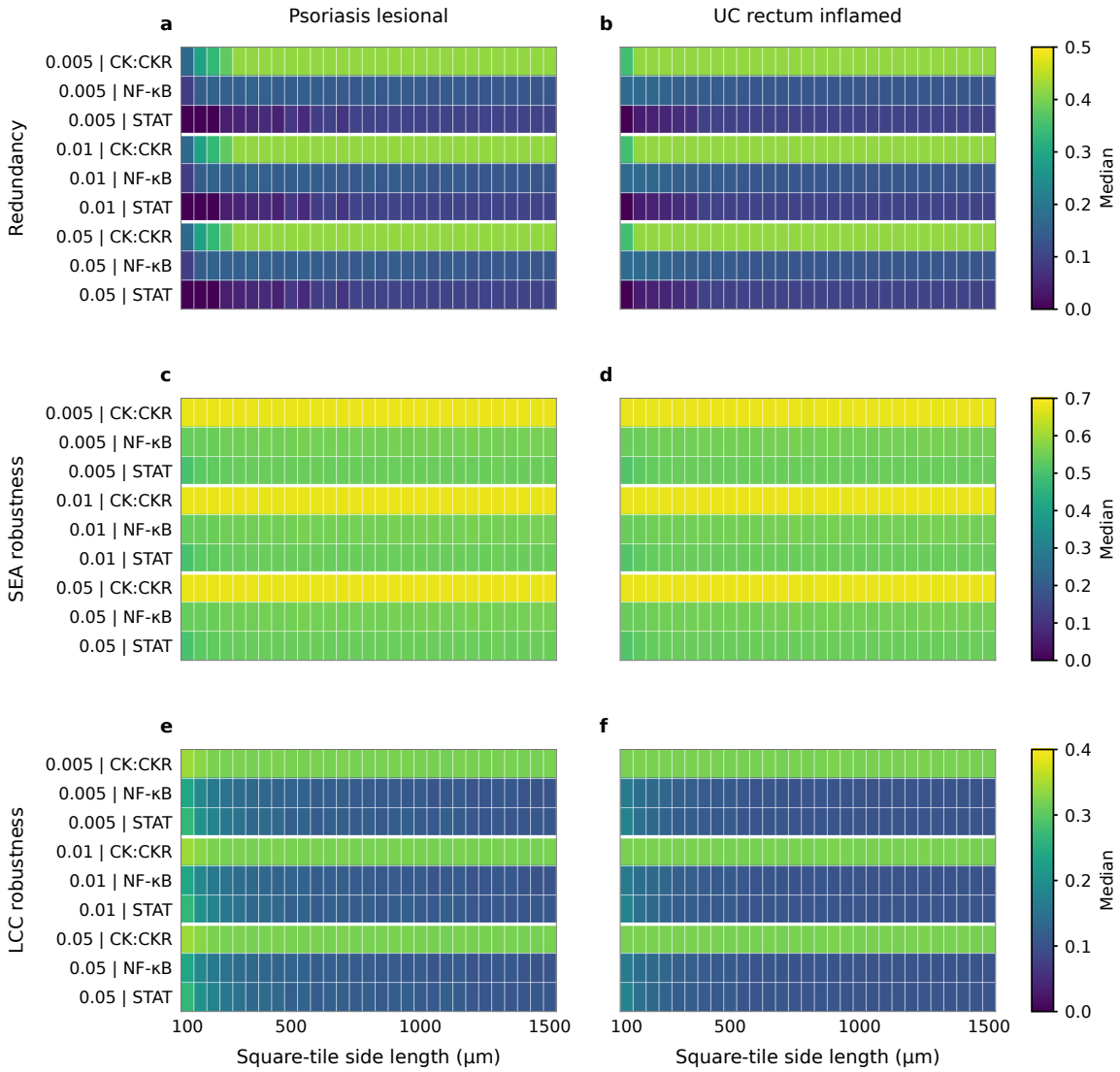

Figure 4.12: Sensitivity of tile-level redundancy and robustness metrics to square-tile side length and TP10K expression threshold. Panels **a** and **b** show median redundancy for psoriasis lesional skin and ulcerative colitis inflamed rectal tissue. Panels **c** and **d** show median secondary extinction AUC for psoriasis lesional skin and ulcerative colitis inflamed rectal tissue. Panels **e** and **f** show median largest connected component AUC for psoriasis lesional skin and ulcerative colitis inflamed rectal tissue. Tile side lengths ranged from 100–1500 μm in 50 μm increments. Expression thresholds were 0.005, 0.01, and 0.05 TP10K.

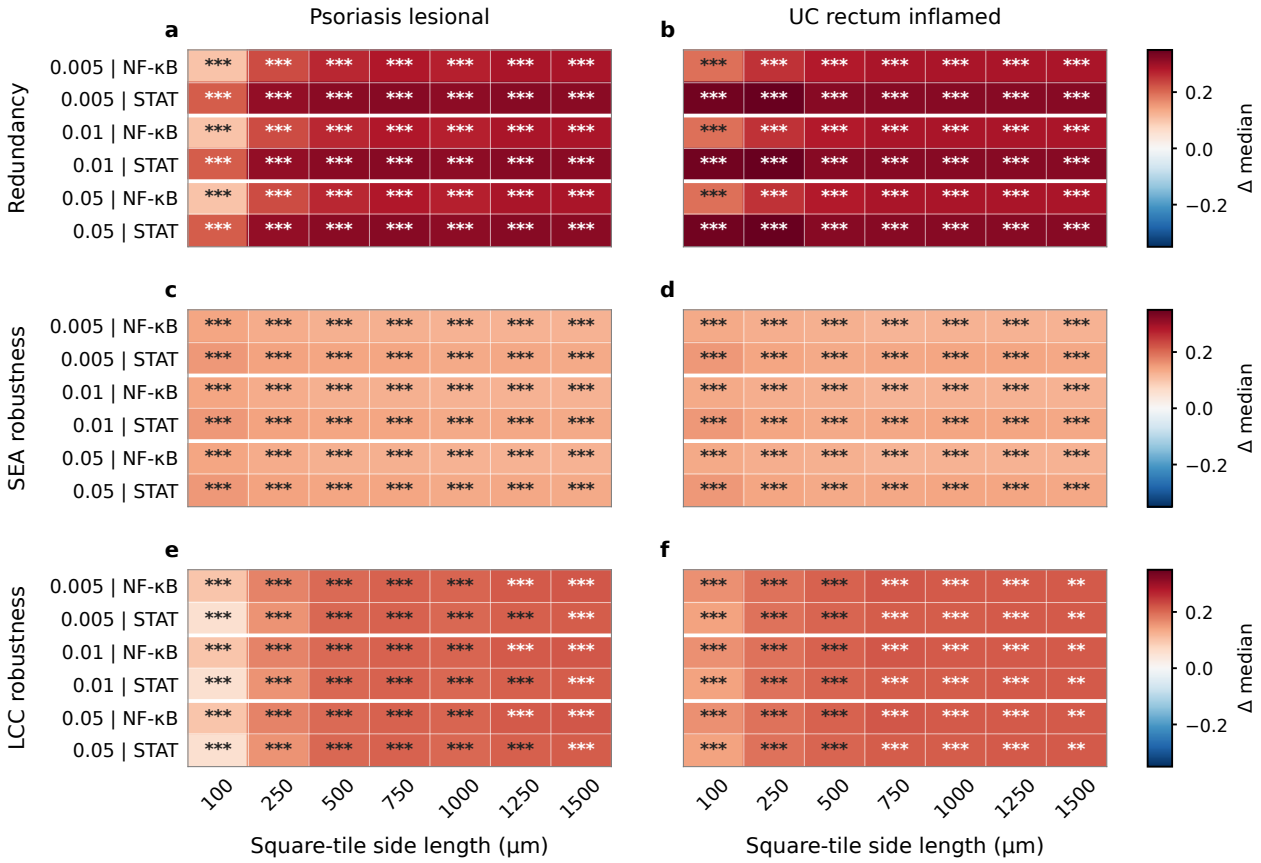

Figure 4.13: Pairwise Wilcoxon signed-rank tests comparing CK:CKR with NF-κB and STAT tile-level network metrics. Tests compared signalling axes across shared tiles, and p values were adjusted with the Benjamini-Hochberg method. Metrics shown are redundancy index, secondary extinction AUC, and largest connected component AUC across tile sizes and TP10K thresholds. Values show the median paired difference (CK:CKR minus comparator). Asterisks indicate adjusted $p \leq 0.05$ (*), $p \leq 0.01$ (**), and $p \leq 0.001$ (***).

4.3 Discussion

This analysis provides the first systematic comparison of cytokine network architecture across the three major signalling axes in human immune-mediated inflammatory disease. The central finding is that CK:CKR networks are characterised by greater edge redundancy and robustness than networks for NF-κB or STAT cytokines. This

difference is consistent across all twelve diseases examined and is robust to variation in expression thresholds used for network construction. The median CK:CKR redundancy index of 0.371 means that CK:CKR networks contained a substantial fraction of edges beyond the minimum needed for node coverage, whereas median NF-κB and STAT redundancy values were much lower.

The consistency across diseases is as important as the size of the effect. The twelve diseases differ in tissue site, dominant pathology, available control samples, and clinical response to existing therapies. Despite this heterogeneity, the CK:CKR axis showed higher redundancy and robustness in every disease-specific comparison. This pattern supports the view that chemokine network architecture is a general property of inflammatory tissue rather than a feature restricted to one disease class. It also reduces the likelihood that the result is explained only by one outlier dataset or one inflammatory condition with exceptionally high chemokine expression.

The spatial single-cell RNA-seq analysis adds a tissue-position requirement to the network result. In psoriasis lesional skin and ulcerative colitis inflamed rectal tissue, CK:CKR networks still had higher redundancy and robustness than NF-κB and STAT networks when a ligand-receptor edge required ligand expression and receptor expression by an immune cell within the same square tissue tile, as shown in **Figures 4.11 and 4.13**. The two spatial datasets therefore support the same ordering after ligand-receptor edges were restricted by tissue position. Both datasets used targeted gene panels, so the spatial result should be interpreted as a focused analysis of two disease tissues rather than a whole-transcriptome spatial survey.

A limitation of this comparison is that NF-κB and STAT group cytokines by downstream pathway, whereas CK:CKR is a ligand–receptor family[59,60,62]. Differences in receptor architecture and interaction-resource coverage may also affect connectivity[137],

potentially limiting generalisation to cytokine signalling as a whole.

All three represent cytokine signalling through cell-surface receptors and were analysed using the same network-construction rules. CK:CKR nevertheless shows consistently greater structural redundancy and robustness than the NF-κB and STAT reference networks across diseases and expression thresholds. This supports a difference in transcript-defined network connectivity. Preserved signalling or migration after ligand blockade still requires functional validation.

The mechanistic basis for elevated CK:CKR network redundancy lies in ligand-receptor promiscuity. Network edge redundancy correlates strongly with promiscuity, particularly receptor-side promiscuity. Chemokine receptors can bind multiple ligands with different affinities and signalling properties[61,78]. In the networks analysed here, blockade of one ligand therefore leaves other receptor connections intact. Under the metric used here, additional ligand-receptor edges increase the number of edges beyond the minimum needed for node coverage when the node set is unchanged. The biochemical promiscuity documented in vitro[61] thus translates directly into network-level redundancy when chemokines and receptors are co-expressed in human disease tissues.

The robustness analyses extend this conclusion from static network topology to simulated ligand-removal behaviour. Secondary extinction analysis, adapted from ecological network studies[124], simulates the consequences of progressive ligand loss: receptors remain active as long as at least one activating ligand persists, but become silent when all their ligands are removed. CK:CKR networks maintain receptor function substantially longer under this regimen than NF-κB or STAT networks. Largest connected component analysis records the fraction of original graph nodes in the largest connected component after each ligand-removal step[125]. CK:CKR networks retain a larger connected component for longer under the same simulation. Both

analyses support the interpretation that chemokine signalling can tolerate perturbation of individual components in ways that NF-κB and STAT signalling cannot.

The use of two robustness measures also limits overinterpretation of any single graph metric. Secondary extinction treats receptor activation as the key endpoint, which is biologically relevant because receptors are the immediate signalling nodes on immune cells. Largest connected component robustness treats the size of the largest connected component, normalised to the original node count, as the endpoint. A result appearing in both measures is therefore more persuasive than one appearing in only one. In this analysis, both measures point in the same direction: CK:CKR networks lose receptor connectivity and largest-component size more slowly under ligand removal.

This result has two clinically relevant interpretations. First, ligand-targeted therapies that neutralise one chemokine are likely to leave other ligands capable of engaging the same receptors. Second, receptor-targeted therapies that inhibit one receptor may still leave ligands free to activate other receptors. The analysis does not specify which exact compensatory pathway dominates in a given patient, but it shows that the expressed disease networks contain the architectural capacity for compensation on both sides of the bipartite graph.

These findings directly address the long-standing debate over redundancy in the chemokine system[87,88]. Proponents of context-specificity have argued that apparent redundancy reflects experimental oversimplification, particularly reliance on inbred mouse models and in vitro assays that fail to capture the complexity of in vivo signalling[88,93,112]. This analysis uses human data from diseased tissues, removing the confounds of species differences and artificial experimental conditions.

The discrepancy between these findings and mouse knockout studies[88,112] may reflect fundamental differences between these experimental contexts[114,115]. Inbred mice have

near-identical inherited genetic backgrounds, reducing inherited genetic variation that might enable compensatory responses in outbred populations[114]. Pathogen-free housing produces naive-skewed immune systems with limited effector and memory T-cell pools, reducing the diversity of chemokine-expressing and chemokine-responsive cells[115]. Humans, in contrast, live in environments with continuous microbial exposure and accumulate diverse immune experience throughout life[115,116]. These conditions may both require and enable chemokine network redundancy: the unpredictable challenges of natural environments favour robust backup systems, while immune experience generates the cellular diversity needed to manifest them[87].

The evolutionary history of the chemokine system provides a mechanistic explanation for why CK:CKR networks differ from NF-κB and STAT networks[80]. The chemokine system originated in early vertebrates approximately 500 million years ago[80]. Rel/NF-κB proteins arose near the origin of opisthokonts, whereas the JAK–STAT pathway formed before chordates[159,160]. Chemokine receptors diversified early through gene duplication in the vertebrate stem group, then remained relatively stable, while chemokine ligands underwent successive expansions in jawed vertebrates, bony fish, and especially placental mammals[80]. This asymmetric expansion generated the ligand-receptor imbalance that underlies promiscuity: with many more ligands than receptors, individual receptors necessarily bind multiple ligands[58,80]. In contrast, cytokines of the NF-κB and STAT axes show more balanced ligand-receptor ratios that have remained stable across vertebrate evolution.

The selective pressures that drove chemokine ligand expansion remain uncertain, but host-pathogen coevolution provides a plausible mechanism[94,95]. Multiple mammalian viruses, including poxviruses, herpesviruses, and cytomegaloviruses, encode chemokine-binding proteins that modulate host immunity[94]. These viral proteins typically bind multiple chemokines, suggesting that effective immune evasion requires broad-spectrum

inhibition[94,95]. The expansion of host chemokines could represent an adaptive response to viral pressure: diversifying the ligand repertoire makes it harder for any single viral protein to neutralise the entire system[94]. This coevolutionary dynamic would naturally generate the redundant network architecture observed in human disease.

Two mechanisms may explain the therapeutic success of NF-κB and STAT pathway inhibitors despite the continued failure of chemokine-targeted approaches[9]. First, the lower redundancy of NF-κB and STAT networks may render them vulnerable to single-target blockade in ways that chemokine networks are not. When TNF is neutralised, there is no equally potent alternative ligand to maintain NF-κB activation through TNFR1; when IL-6 is blocked, the IL-6 family has limited capacity for compensation. Second, NF-κB and STAT pathways lie upstream of chemokine production: inflammatory stimuli activate these transcription factors, which then induce chemokine gene expression[63,64]. Drugs that block NF-κB or STAT signalling therefore suppress chemokine production indirectly, achieving multi-chemokine inhibition through a single molecular target. This indirect mechanism may explain why TNF inhibitors reduce leukocyte infiltration in rheumatoid synovium and inflammatory bowel mucosa despite not directly targeting chemokine signalling[11,28].

The therapeutic implications of these findings are direct[86]. If chemokine network redundancy underlies the failure of single-target inhibitors, then effective therapeutic intervention may require multi-target approaches[86,95]. The convergent evolution of multi-chemokine inhibitors in ticks and viruses supports this view: these organisms have independently evolved ways to disable a redundant network, and their solutions involve broad-spectrum binding rather than target specificity[6,94]. Understanding how these natural inhibitors achieve multi-chemokine binding may inform the design of therapeutic agents capable of overcoming network robustness.

One counterargument to this conclusion might be that it requires every chemokine interaction to be biologically interchangeable. However, it does not. The network results are compatible with ligand-receptor pairs having different potencies, kinetics, tissue distributions, or biased signalling properties. The point is narrower: in diseased human tissues, the expressed chemokine system contains many alternative ligand-receptor connections, and these alternatives preserve receptor connectivity under simulated perturbation. Context-specificity may therefore coexist with redundancy. Individual pathways may have specialised functions, while the overall network remains difficult to disable through single-target intervention.

The next chapter asks how such a network might be attacked successfully. If robustness arises from many overlapping chemokine inputs, then an effective inhibitor should neutralise multiple chemokines or disrupt shared chemokine functions. Tick evasins provide a natural example of this strategy. They do not inhibit a single receptor in the manner of conventional receptor antagonists; instead, they bind multiple chemokine ligands. The structural analysis that follows examines how evasins achieve this broad binding and whether they target chemokine surfaces required for receptor interaction and self-association.

Chapter 5

Characterisation and Structural Analysis of the Ability of Tick Evasins to Attack Chemokine Networks in Inflammatory Disease

5.1 Introduction

The network analysis in Chapter 4 showed that human disease CK:CKR networks have greater redundancy and robustness than NF-κB and STAT networks. This property provides a mechanistic explanation for the failure of single-target chemokine inhibitors in clinical trials[7,8]. The therapeutic question that follows is how chemokine network robustness can be overcome. Nature provides an answer: blood-feeding parasites have independently evolved proteins that bind and neutralise multiple chemokines simultaneously[6,94,95].

Ticks are obligate blood-feeding arachnids that require prolonged attachment to their host, typically lasting several days[6,126]. During this period, the host mounts an inflammatory response directed at the feeding site, recruiting neutrophils, macrophages,

and lymphocytes that threaten to interrupt attachment or impair blood acquisition[6]. To counteract this response, ticks secrete a complex cocktail of salivary proteins with immunomodulatory properties[6,126]. Among these are the evasins, a family of chemokine-binding proteins identified in multiple tick genera[6,127,128].

Evasins fall into two structural classes with distinct chemokine specificities[6,126]. Class A evasins contain four to five disulfide bonds and bind CC chemokines; Evasin-1 and Evasin-4 from *Rhipicephalus sanguineus* are prototypic Class A evasins. Evasin-1 and the Class B protein Evasin-3 were the first tick chemokine-binding proteins to be characterised structurally[129]. Class B evasins adopt a knottin scaffold with three disulfide bonds and bind CXC chemokines[6,131]. Both classes bind multiple chemokines within their respective subfamilies. Evasin-1 binds six reported CC chemokines, whereas Evasin-4 binds approximately 20 CC chemokines[6].

The therapeutic relevance of multi-chemokine binding has been demonstrated in animal models[95,103,129]. Evasin-1 suppresses inflammation in mouse models of peritonitis and lung injury[129]. It also reduces mortality following intestinal ischaemia-reperfusion[129]. Evasin-4 reduces post-infarction myocardial injury[103]. It also reduces disease severity in models of contact hypersensitivity and colitis[103]. These efficacy signals contrast sharply with the clinical failure of single-target antagonists and support the hypothesis that network redundancy must be overcome for effective therapeutic intervention[86,95]. Understanding the structural mechanisms by which evasins may achieve broad-spectrum binding could therefore inform the design of next-generation therapeutics[86,132].

How does a single protein bind diverse ligands that share only limited sequence identity? Two broad mechanisms are conceivable. In the conformational plasticity model, the evasin binding interface adopts different conformations to accommodate the distinct surface properties of different chemokine partners. The binding site itself is flexible,

reshaping to maintain complementarity despite sequence variation among targets. In the differential positioning model, the evasin presents distinct surface patches that engage different chemokines. Broad specificity arises from having multiple binding sites rather than from flexibility of a single site. These mechanisms have different implications for therapeutic design: conformational plasticity could potentially be engineered into synthetic scaffolds, whereas differential positioning might require larger, multi-domain architectures.

Bhusal et al. resolved EVA-P974 bound to CCL7 and CCL17, a CCL8-CCL7 chimera, and an engineered EVA-P974 variant bound to CCL2[161]. Devkota et al. subsequently determined structures of the class A3 evasin EVA-AAM1001 bound to CCL7, CCL11, CCL16 and CCL17, together with engineered complexes[162]. A later study reported EVA-ACA1001 bound to CCL16[163]. These studies provide direct examples of recognition across several chemokine partners and of structure-guided changes in target preference[161,162]. Experimental coverage nevertheless remains incomplete across the full set of evasin-chemokine interactions, supporting the use of prediction to extend these structural observations. AlphaFold 3, released in 2024, offers a computational approach to this problem[146]. AlphaFold 3 predicts the structures of protein-protein complexes from sequence alone, achieving near-experimental accuracy for many interaction types[146]. By modelling the same evasin in complex with different chemokine partners, it is possible to assess structural variation at the binding interface and distinguish between the conformational plasticity and differential positioning models.

This chapter describes the structural analysis of evasin-chemokine interactions using AlphaFold 3. I compile experimentally validated evasin-chemokine interactions from the literature, generate structural models of these complexes, validate predictions against available experimental structures, and analyse the binding interface to determine whether evasins may achieve broad specificity through conformational plasticity,

differential positioning, or some combination. I further examine how predicted evasin binding relates to the functional surfaces of chemokines used for receptor interaction and self-association, testing whether evasins may act by sterically occluding these sites. The goal is to identify predicted structural principles that suggest how natural inhibitors attack a redundant chemokine network.

5.2 Results

5.2.1 Compilation of chemokine interaction data

The workflow is summarised in **Figure 5.1**. It integrates literature-derived interaction data with structure prediction and binding-interface analysis.

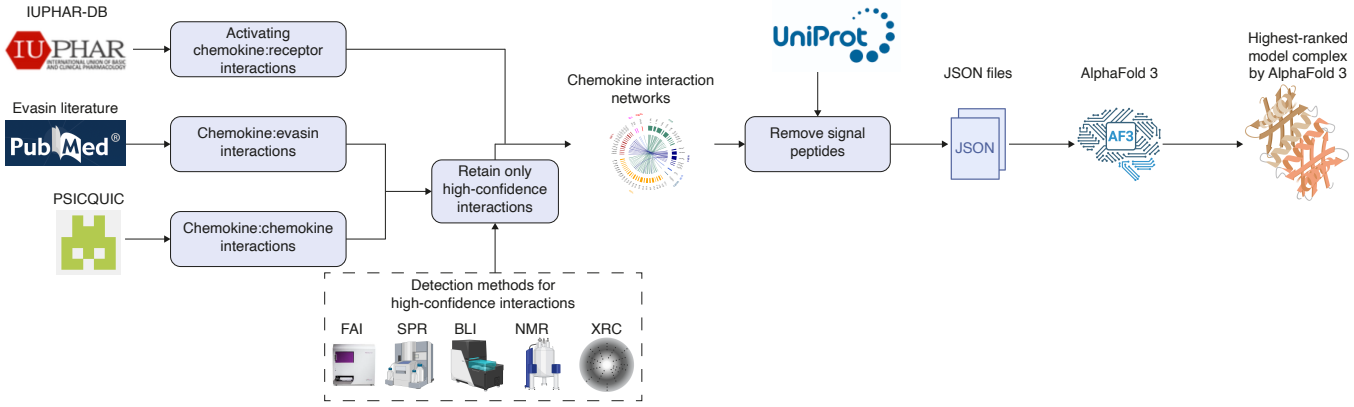

Figure 5.1: Flow diagram describing the computational pipeline for identifying and structurally characterising the chemokine network. Chemokine network members were identified by searching IUPHAR-DB for activating chemokine:chemokine receptor interactions, literature searches for chemokine:evasin interactions, and PSICQUIC for chemokine:chemokine interactions. Network chord diagrams were constructed using high-confidence interactions. Structural characterisation of complexes was performed by downloading protein sequences of chemokine network members from UniProt, removing the signal peptide where present, and generating JSON files that were submitted to the AlphaFold 3 webserver. The highest-ranked model for each complex was used for all subsequent analyses.

Chemokine interaction data were compiled from multiple sources. Evasin-chemokine in-

teractions totalling 386 pairs were identified from the literature[6,126,131–133,155]. Chemokine-chemokine interactions totalling 71 pairs were identified from PSICQUIC. Activating chemokine-receptor interactions totalling 66 pairs were identified from IUPHAR-DB[61].

Interactions measured by biolayer interferometry, surface plasmon resonance, NMR spectroscopy, or X-ray crystallography were included in structural modelling. After this evidence filter, 523 high-confidence interactions remained. Chord diagrams of the chemokine:evasin, chemokine:chemokine receptor, and chemokine:chemokine interaction sets are shown in **Figures 5.2 to 5.4**.



Figure 5.2: Chord diagram of chemokine:tick evasin interactions used for structural modelling. Nodes arranged around the circumference represent chemokines and tick evasins, and chords connect experimentally supported chemokine:evasin binding pairs. CCL chemokines are shown in orange, CXCL chemokines in red, and evasins in green. The interaction set contains 386 literature-derived chemokine:evasin pairs supported by physical binding evidence and was used to define AlphaFold 3 modelling targets[6,126,131–133,155].

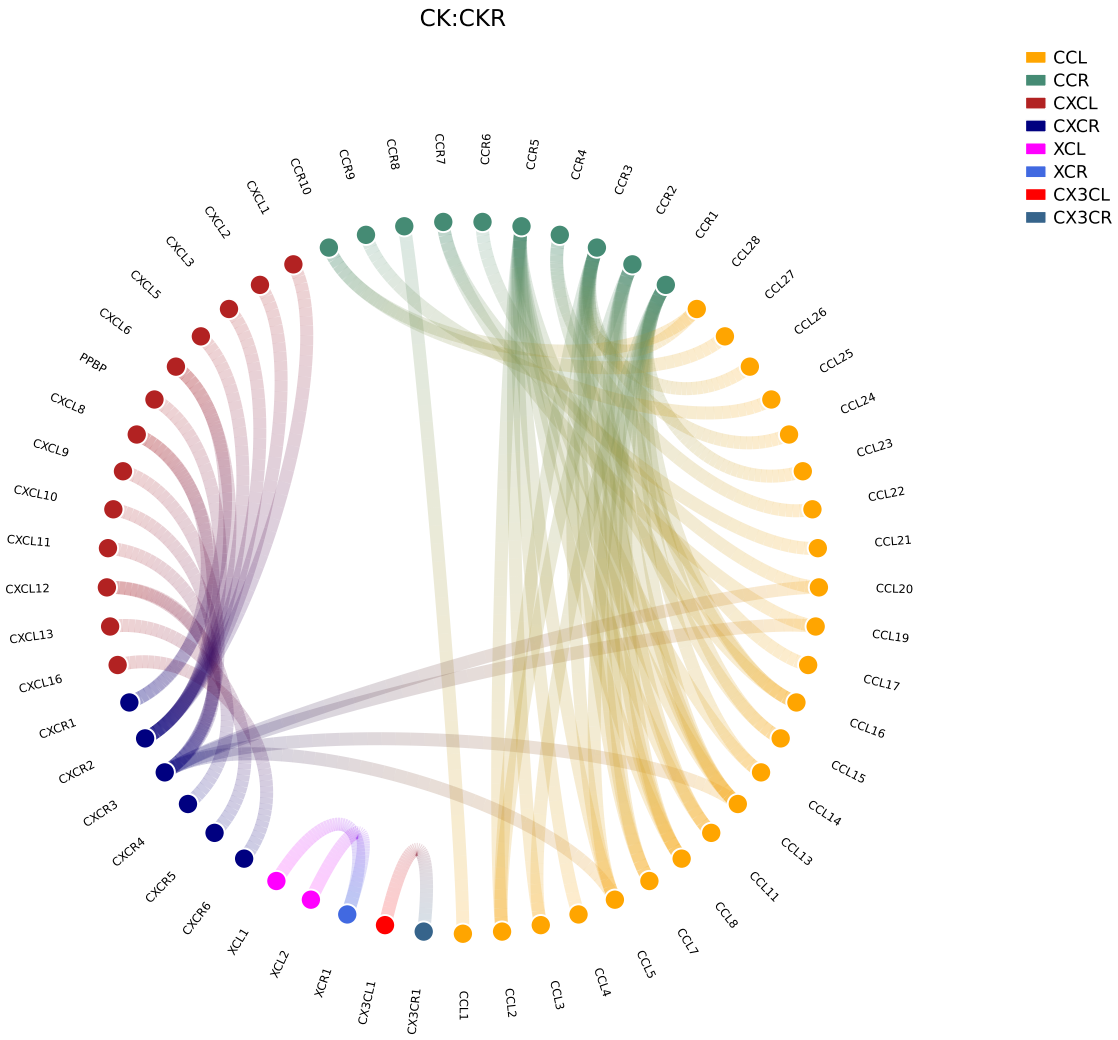

Figure 5.3: Chord diagram of activating chemokine:chemokine receptor interactions used for structural modelling. Nodes arranged around the circumference represent chemokines and chemokine receptors, and chords connect curated activating ligand-receptor pairs. Chemokine and receptor families are indicated by colour, with CCL and CCR families, CXCL and CXCR families, XCL and XCR, and CX3CL and CX3CR shown separately. The interaction set contains 66 activating chemokine-receptor pairs identified from IUPHAR-DB and was used to define AlphaFold 3 modelling targets[61].

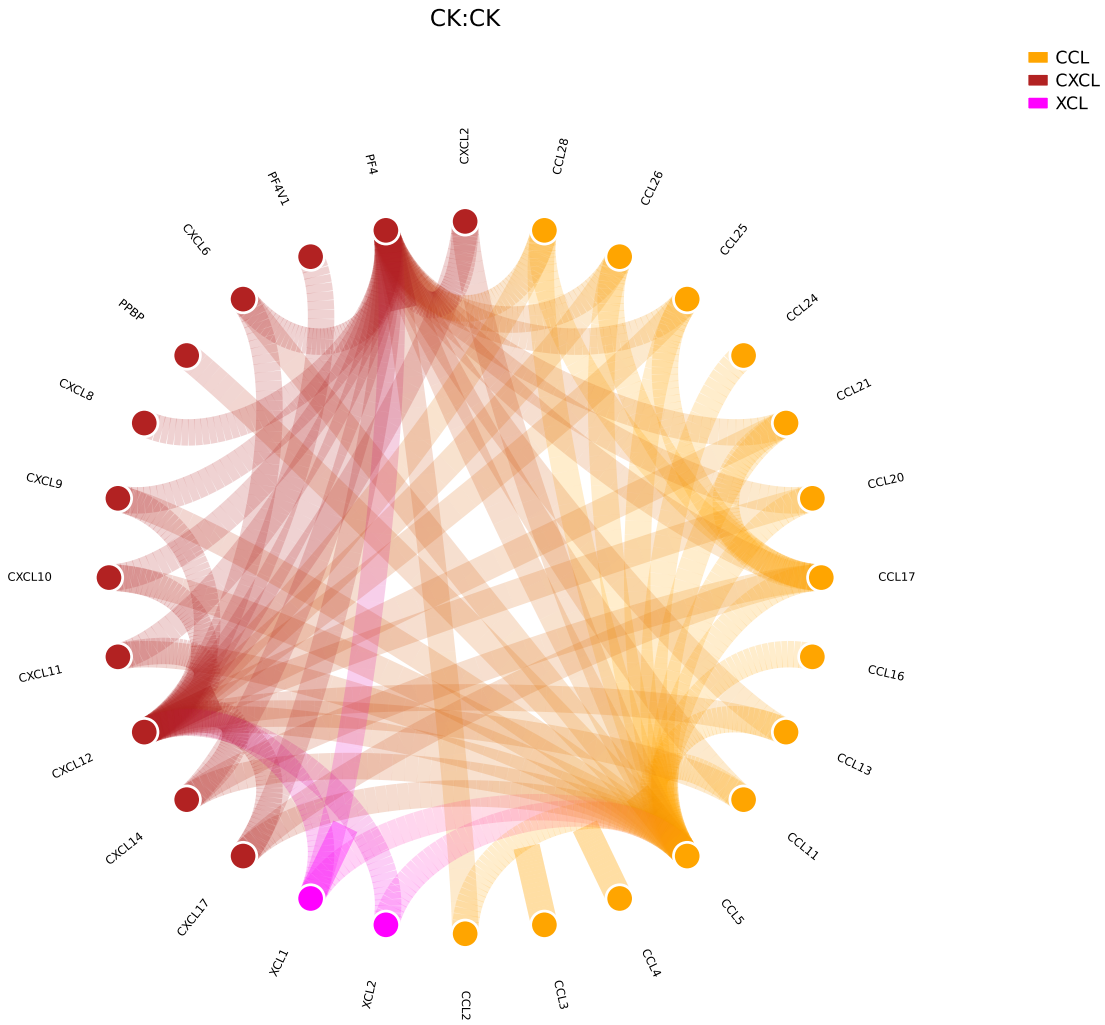

Figure 5.4: Chemokine–chemokine interaction network used for structural modelling. Each node represents a chemokine included in the high-confidence chemokine–chemokine interaction set, and curved links connect chemokine pairs with experimentally supported physical interactions. Node colour indicates chemokine subfamily: CCL chemokines are shown in orange, CXCL chemokines in red, and XCL chemokines in magenta. This interaction set was used to select chemokine dimer models for downstream contact and binding-site overlap analyses.

This interaction set deliberately prioritised experimentally supported physical binding over predicted or text-mined associations. That decision was important because the structural analysis depends on the existence of a plausible protein-protein complex for each pair. Including unsupported associations would increase coverage but would also risk modelling complexes that do not form in reality. The final set therefore represents a conservative substrate for AlphaFold 3 modelling: it is smaller than the universe of possible chemokine interactions, but each included pair has experimental support for physical interaction or receptor activation.

The three interaction classes serve different biological roles. Chemokine:receptor interactions define the activating edges that drive leukocyte responses. Chemokine:chemokine interactions represent self-association or heteromeric association, which can shape gradient formation, receptor activation, and ligand availability[57,84]. Chemokine:evasin interactions represent natural inhibitory relationships evolved by ticks. Analysing all three classes in the same structural framework allowed evasin binding sites to be compared directly with the functional surfaces used by chemokines for signalling and self-association.

5.2.2 AlphaFold 3 structure prediction

For the 523 high-confidence interactions, models of the interacting partners were generated using AlphaFold 3 and the top-ranked prediction per pair was analysed[146]. Representative AlphaFold 3 models of chemokine:evasin, chemokine:chemokine receptor, and chemokine:chemokine complexes are shown in **Figures 5.5 to 5.7**.

Only the highest-ranked model for each pair was carried forward, so the analysis should be interpreted as a systematic comparison of predicted dominant binding poses rather than an ensemble analysis. This is appropriate for asking whether predicted interfaces overlap across many complexes, but it does not capture all possible alternate

conformations or transient encounter states. The same rule was applied to every interaction class, which keeps the comparison internally consistent.

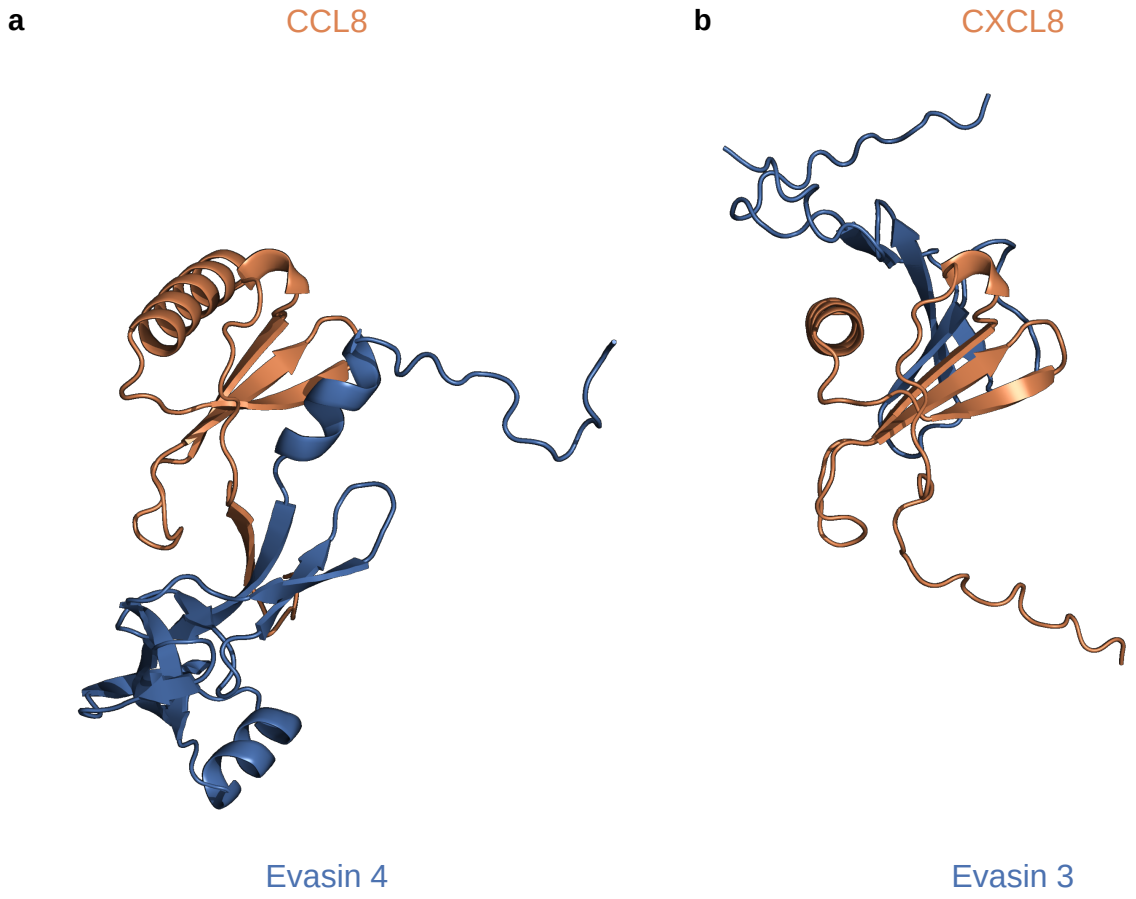

Figure 5.5: Representative AlphaFold 3 models of chemokine–tick evasin complexes. Panel **a** shows the highest-ranked predicted complex for CCL8 and Evasin 4 (class A), rendered as a protein cartoon. Panel **b** shows the corresponding model for CXCL8 and Evasin 3 (class B). Chemokines share a common orientation to compare the regions bound by each evasin. The models illustrate the structural format used for downstream analysis of evasin-contact residues, chemokine binding-site overlap, and predicted steric occlusion of functional chemokine surfaces.

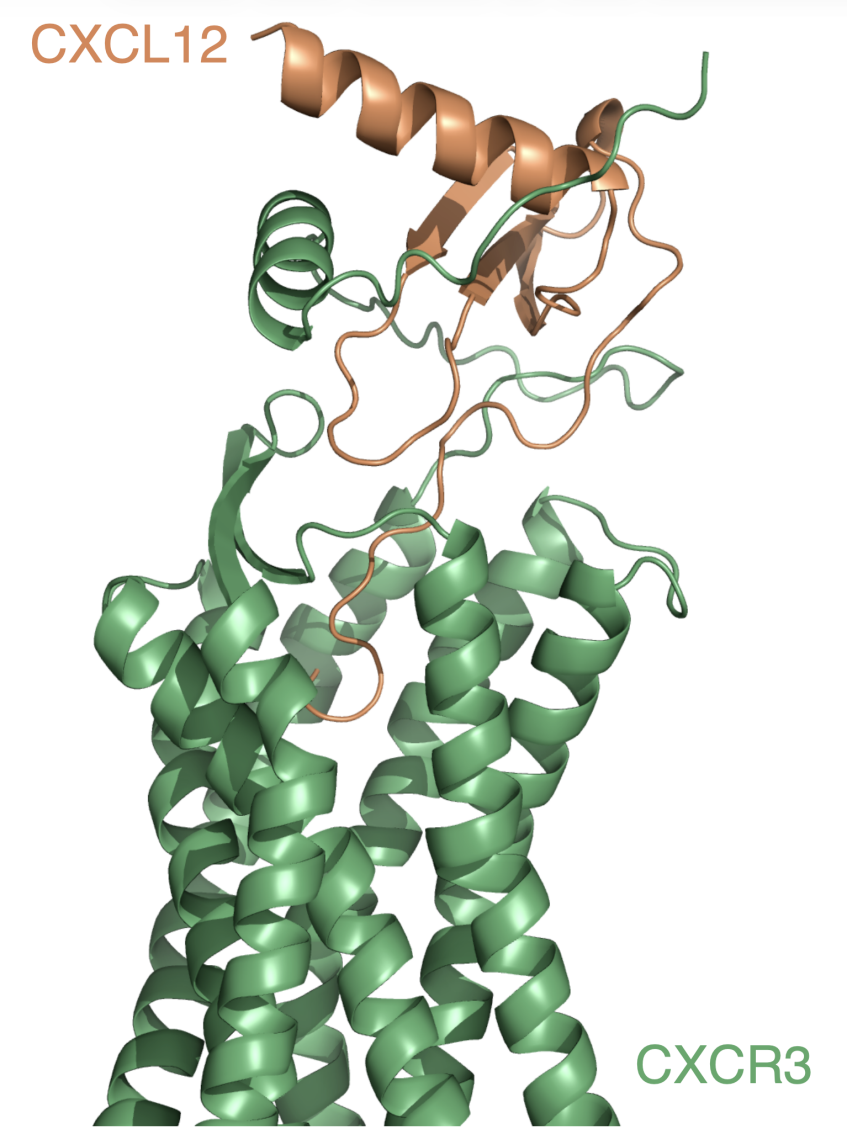

Figure 5.6: Representative AlphaFold 3 model of a chemokine–chemokine receptor complex. The example shows the highest-ranked predicted complex for CXCL12 and CXCR3, rendered as a protein cartoon. CXCL12 is shown in orange and the chemokine receptor is shown in green. The model illustrates the structural format used for downstream analysis of chemokine receptor-contact residues and comparison with evasin-binding surfaces.

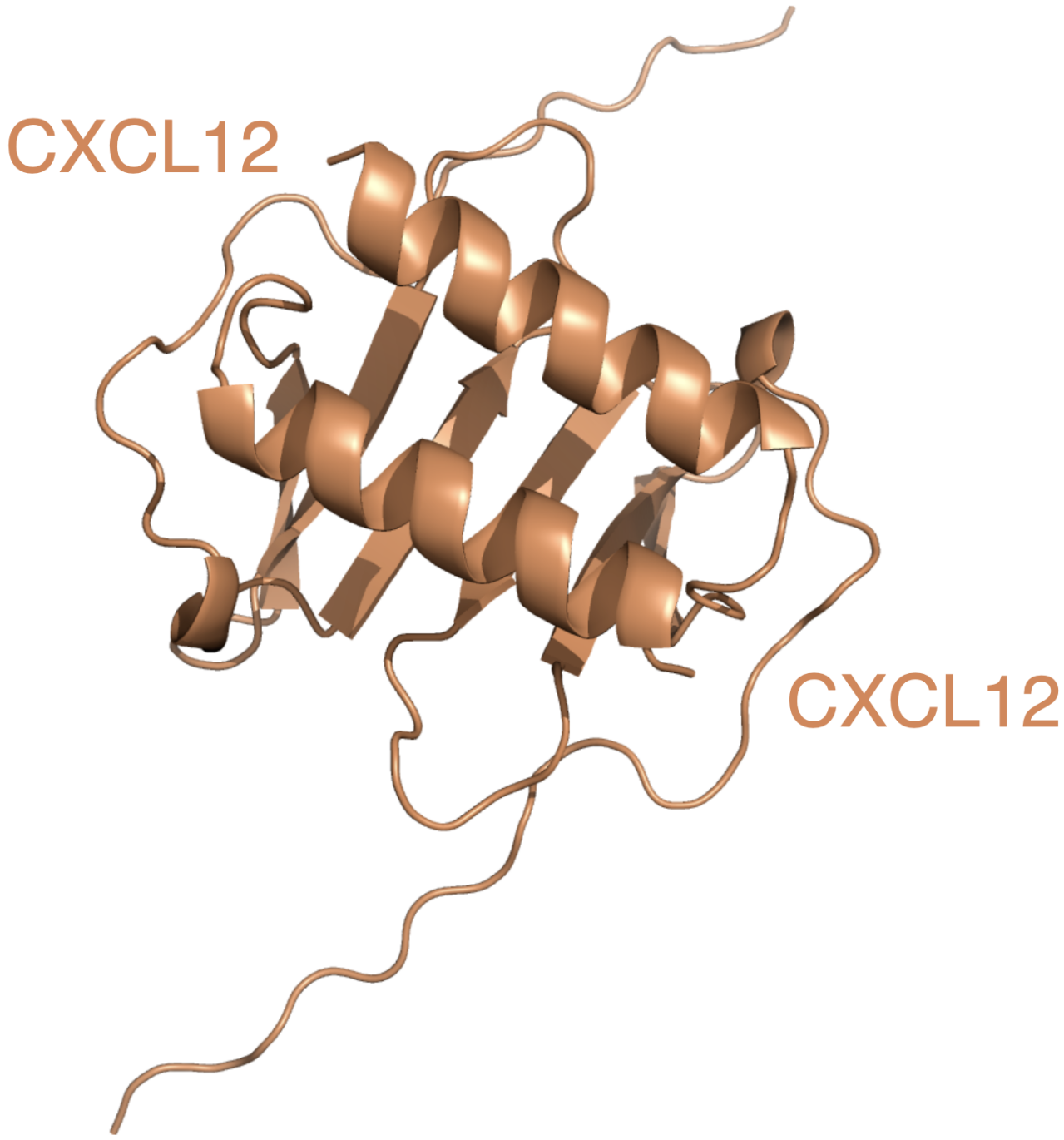

Figure 5.7: Representative AlphaFold 3 model of a chemokine–chemokine complex. The example shows the highest-ranked predicted CXCL12 homodimer model, rendered as a protein cartoon with both chemokine chains shown in orange. The model illustrates the structural format used for downstream analysis of chemokine dimer contacts and binding-site overlap.

5.2.3 Validation against experimental structures

Concordance with experimentally determined PDB structures was assessed by root mean square deviation. Good agreement, defined as RMSD below 2.5 Å, was observed in 41 of 42 cases where experimental structures were available. The distribution of post-alignment RMSD values between AlphaFold 3 models and corresponding PDB structures is shown in **Figure 5.8**.

This validation step supports the use of the models for comparative interface analysis, but it does not mean every predicted complex is experimentally proven. The PDB validation set covers interactions for which structures already exist, whereas many evasin-chemokine pairs lack experimental structures. The result therefore demonstrates that AlphaFold 3 reproduced available chemokine complex structures in almost all testable cases, providing confidence in the modelling approach for this protein family.

To test predictions outside PDB coverage, the AlphaFold 3 model of a class B evasin-chemokine complex, specifically EVA3-CXCL8, was compared with an NMR-based EVA3-CXCL8 model. Using the backbone-only align command in PyMOL, agreement was found to be high with an RMSD of 2.803 Å, and the predicted binding interface matched the NMR-based model. As no class B evasin-chemokine structures are deposited in PDB, these results support the use of AlphaFold 3 models to examine predicted binding poses for novel evasin-chemokine complexes.

Agreement with an independently generated NMR-based model does not replace experimental structure determination, but it provides an additional check outside the PDB-derived validation set. Together, the PDB comparison and the EVA3-CXCL8 comparison provide justification for proceeding to broad interface-level analyses while retaining the interpretation that the structures are computational predictions.

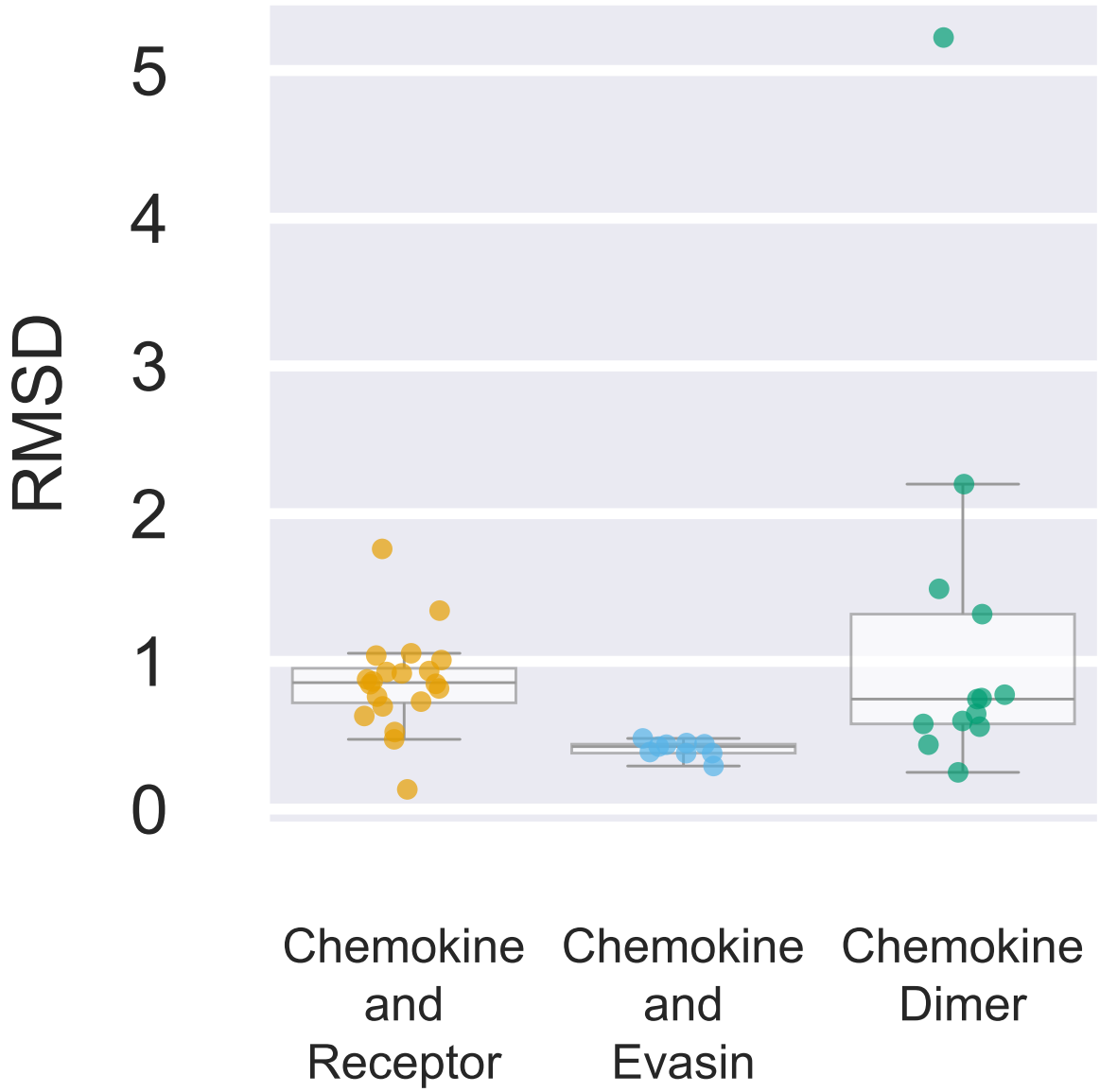

Figure 5.8: Validation of AlphaFold 3 chemokine complex models against experimentally determined structures. Post-alignment root mean square deviation (RMSD) values are shown for chemokine–receptor, chemokine–evasin, and chemokine dimer complexes with corresponding PDB structures. Lower RMSD values indicate closer structural agreement between the predicted model and the experimental structure. Boxes show the median and interquartile range, whiskers show the 1.5 interquartile range interval, and coloured points show individual complex comparisons.

5.2.4 Evasins sterically occlude functional chemokine sites

The mechanism by which evasins inhibit the chemokine network was next examined by comparing evasin-binding surfaces with the chemokine regions used for receptor interaction and homo- or heterodimerisation[57,84]. If an evasin occupies the same surface that a chemokine uses to bind a receptor, receptor activation can be blocked directly. If the evasin occupies a surface used for chemokine dimerisation or heteromer formation, it can also interfere with higher-order chemokine organisation. These two mechanisms are not mutually exclusive; a single evasin-binding surface can overlap both receptor-facing and dimerisation-facing regions on the chemokine.

5.2.5 Systematic analysis of binding site overlap

To systematically analyse binding site overlap across all chemokine interactions, the extent of overlap across all chemokine-chemokine, chemokine-receptor, and chemokine-evasin pairs was examined. This analysis was based on residue-contact frequencies from independently modelled pairwise complexes and was intended to identify recurrently used chemokine surfaces and potential steric competition between partners.

A multiple sequence alignment was first generated for all CC and CXC chemokines that are known to interact with an evasin. For each residue position within the alignment, the proportion of complexes in which that residue was contacted by an interacting partner was plotted. Hotspots for interactions were defined as residues where the proportion of complexes contacted exceeded 0.3. The alignment used to define shared residue positions is shown in **Figure 5.9**.

Examination of contact profiles suggested that chemokines tend to bind their interacting partners through regions that overlap with each other. These regions include the N-terminus and β-sheet in the case of CC chemokines, and the N-terminus, β-sheet,

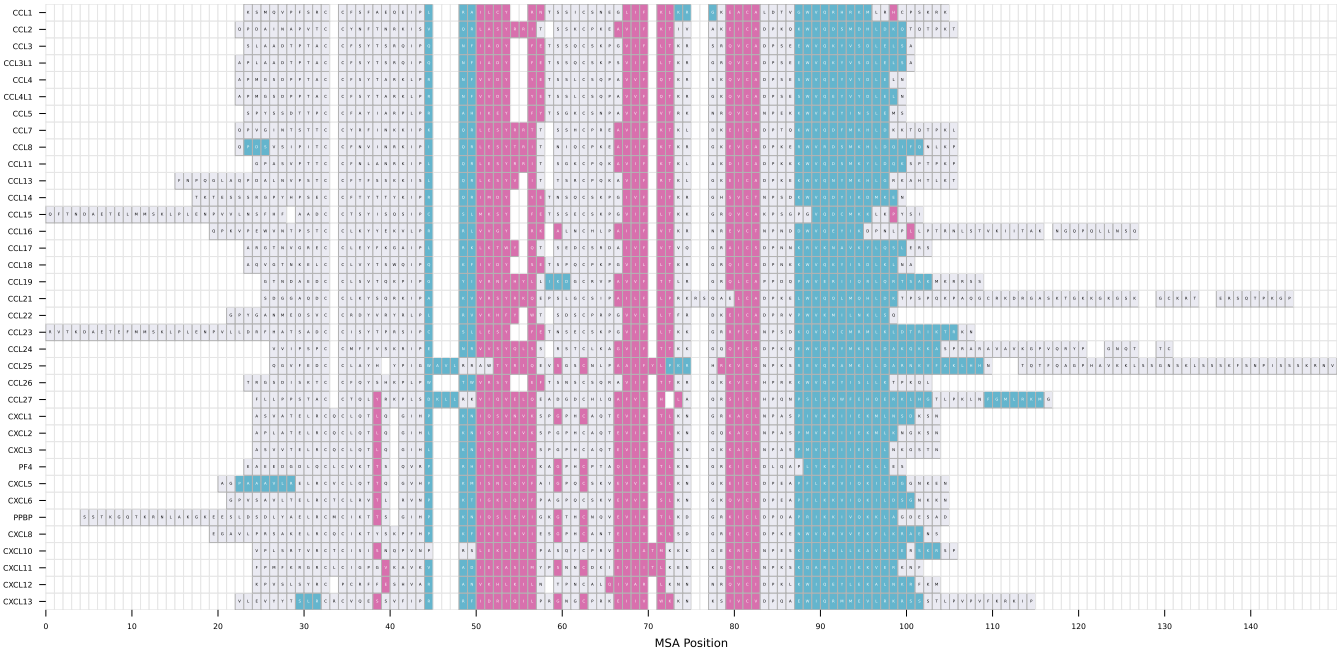

Figure 5.9: Multiple sequence alignment of CC and CXC chemokines used for contact-profile analysis. Rows show individual chemokines and columns show aligned residue positions. Amino-acid identities are shown within occupied cells, with alignment gaps left blank. Secondary structure was assigned from representative AlphaFold 3 complex CIF files by reading PyMOL chain-A Cα annotations, with assignments matched to the chemokine symbols shown on the y-axis before plotting. α-helical positions are shown in blue, β-strand positions in pink, and loop or unstructured positions in light grey. The alignment provides the shared residue coordinate system used to map chemokine contact frequencies from modelled chemokine–chemokine, chemokine–receptor, and chemokine–evasin complexes onto homologous positions.

and C-terminal helix in the case of CXC chemokines. The contact-frequency profiles are shown in **Figure 5.10**.

Because the contact frequencies were calculated across independently modelled complexes, recurrent hotspots indicate surfaces used repeatedly by different interaction partners. At the structural level, this mirrors the network-level observation that many interactions route through shared chemokine features: different partners repeatedly use a limited set of chemokine surfaces. For an inhibitor, this is advantageous because binding to a shared hotspot may block several biological functions at once.

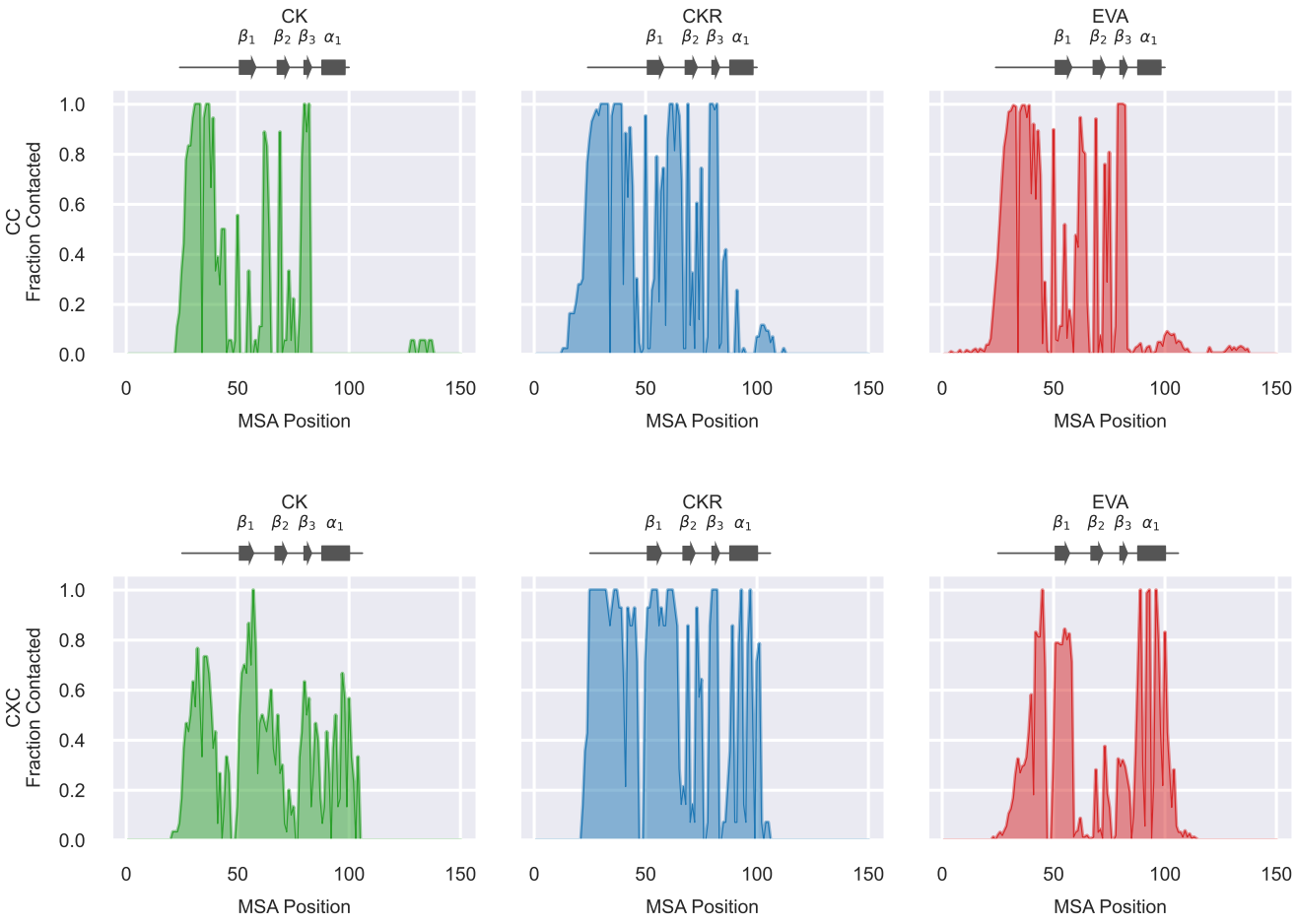

Figure 5.10: Chemokine contact profiles across aligned CC and CXC chemokines. Rows separate CC and CXC chemokines, and columns show chemokine–chemokine interactions (CK), chemokine–receptor interactions (CK:CKR), and chemokine–evasin interactions (EVA). The x-axis shows multiple-sequence-alignment position, and the y-axis shows the fraction of modelled complexes in which each aligned residue position was contacted. Green, blue, and red profiles correspond to CK, CK:CKR, and EVA contact sets, respectively. Schematics show β_1–β_3 strands and the α_1 helix of representative CCL5 (CC) and CXCL10 (CXC) chemokines, mapped to the alignment.

5.2.6 Quantification of binding site overlap

The overlap was tested formally by computing the Jaccard index, which measures how much two sets overlap relative to their total combined size[156]. The Jaccard index ranges from 0, indicating no overlap, to 1, indicating perfect overlap.

For CC chemokines, the Jaccard indices comparing class A evasin binding sites to chemokine receptor-binding sites and to chemokine dimerisation sites were 0.71 and 0.88, respectively. For CXC chemokines, the Jaccard indices comparing class B evasin binding sites to chemokine receptor-binding sites and to chemokine dimerisation sites were 0.54 and 0.37, respectively.

In 3,000 random permutations for each comparison, none produced an overlap as large as the observed value (empirical $p < 3.3 \times 10^{-4}$ for each comparison), indicating that the observed overlaps are far greater than expected by chance. These results confirm substantial overlap in chemokine regions that are used to interact with different partners. These model-derived overlaps suggest that evasins act, at least in part, by sterically occluding chemokine regions used for receptor interaction and chemokine self-association, consistent with competition at shared surfaces.

The Jaccard analysis converts the visual impression of overlap into a reproducible residue-set comparison. High overlap with receptor-binding sites supports predicted direct inhibition of receptor engagement. High overlap with dimerisation sites supports predicted interference with chemokine self-association or heteromer formation. The CC and CXC values differ in magnitude, which is expected because the two chemokine classes differ in structure, dimerisation geometry, and evasin class. The common result is that evasin-binding sites are not randomly distributed across chemokine surfaces; they are concentrated on regions already used for chemokine function.

5.2.7 Conformational plasticity at evasin binding interfaces

To investigate how evasins bind multiple chemokines, two mechanisms were considered. Conformational structural plasticity involves a single binding site on the evasin adopting multiple conformations to accommodate different partners. Differential positioning involves different evasin surface regions engaging different chemokines. The interface

analysis workflow is summarised in **Figure 5.11**.

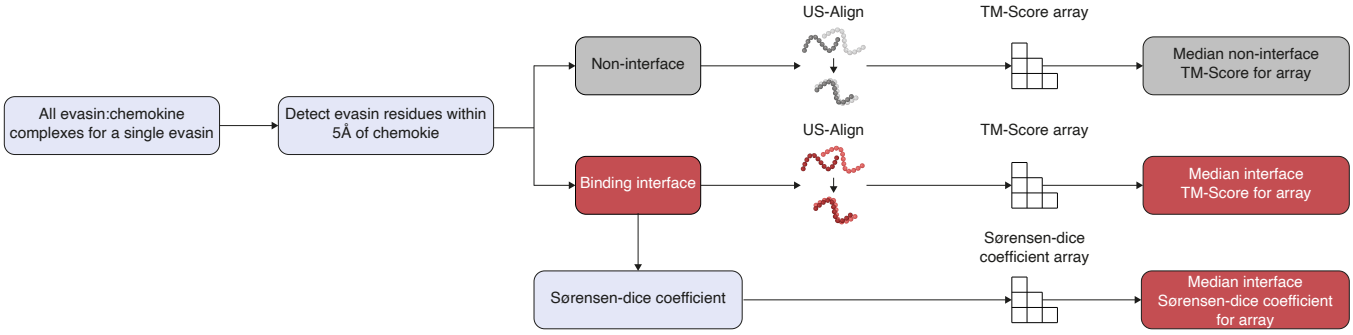

Figure 5.11: Pipeline for evasin interface analysis. For each evasin, all modelled evasin–chemokine complexes were analysed to identify evasin residues within 5 Å of the chemokine partner. Non-interface and binding-interface regions were compared separately using US-align and summarised as pairwise TM-score matrices. Binding-interface residue sets were also compared using Sørensen–Dice coefficients. Each array was summarised by its median pairwise value.

Visual inspection suggested conformational plasticity. For example, EVA1 interface regions can be unstructured in some complexes but form helical turns in others. This was quantified by comparing structural similarity of interface residues and non-interface residues for each evasin bound to different chemokines. The superposed EVA1 interfaces are shown in **Figure 5.12**.

Template modelling scores were used to compare interfaces of different lengths; values closer to 1 denote greater structural similarity and therefore less plasticity[157]. For each evasin with at least two bound chemokines, pairwise TM-score matrices for non-interface regions and interfaces were computed, and each matrix was summarised by its median pairwise TM-score. Representative pairwise matrices are shown in **Figure 5.13**.

Median interface TM-scores were lower than non-interface TM-scores for both class A (0.408 versus 0.764) and class B evasins (0.392 versus 0.606). This suggests greater conformational plasticity at the binding interface compared to the rest of the protein.

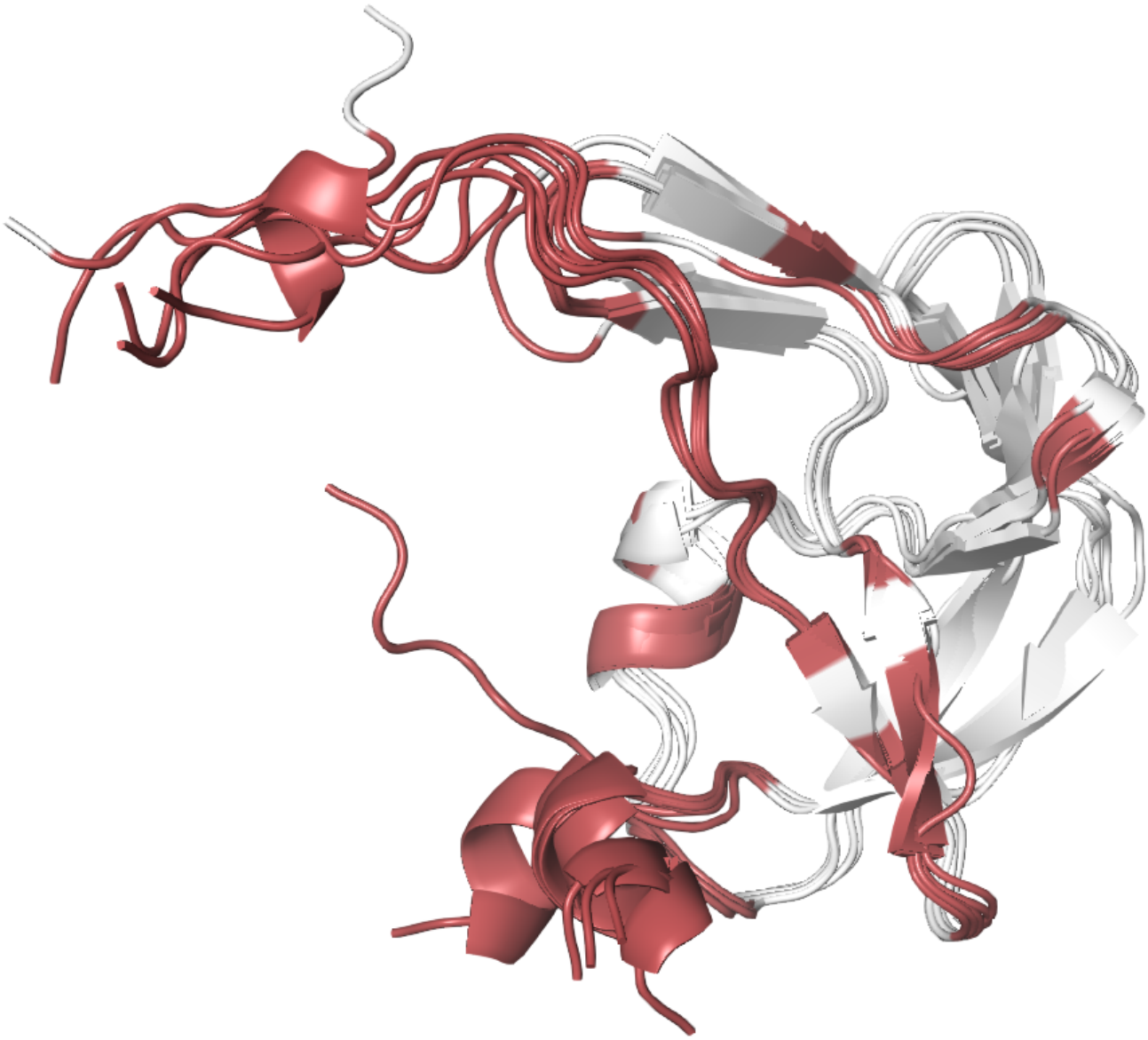

Figure 5.12: EVA1 reuses a shared chemokine-binding surface across modelled complexes. Superposed EVA1 structures from AlphaFold 3 models with different chemokine partners are shown as ribbon traces. Non-interface regions are shown in grey, and chemokine-binding interface residues are shown in red. Interface residues were defined as evasin residues within 5 Å of the chemokine partner. The overlay shows that contacts localise to the same broad surface across complexes, while interface loops and secondary-structure elements adopt different conformations, consistent with local conformational plasticity at a shared binding interface.

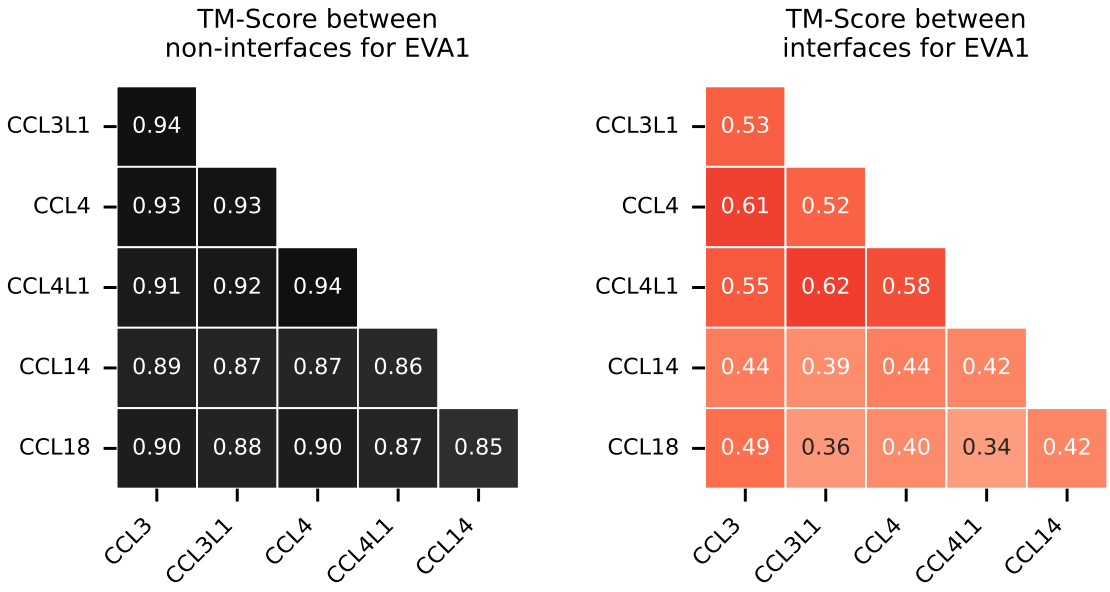

Figure 5.13: Pairwise TM-score matrices for EVA1 non-interface regions (left) and binding interfaces (right) across chemokine partners. Higher TM-scores indicate greater structural similarity. Non-interface regions show consistently high similarity, whereas binding interfaces show lower pairwise similarity, consistent with local conformational plasticity at the binding surface.

The binding sites of evasins are predicted to be more structurally variable than their core regions. The class-level distributions are shown in **Figure 5.14**.

This result supports the predicted conformational plasticity model because the interface changes more than the rest of the evasin. If the entire evasin changed equally between complexes, the result would indicate global structural instability rather than local binding-site adaptation. If the interface remained as rigid as the non-interface region, broad specificity would need to be explained by a different mechanism. The observed pattern is intermediate and biologically plausible: a conserved structural scaffold presents a binding region that can reshape to accommodate different chemokine partners.

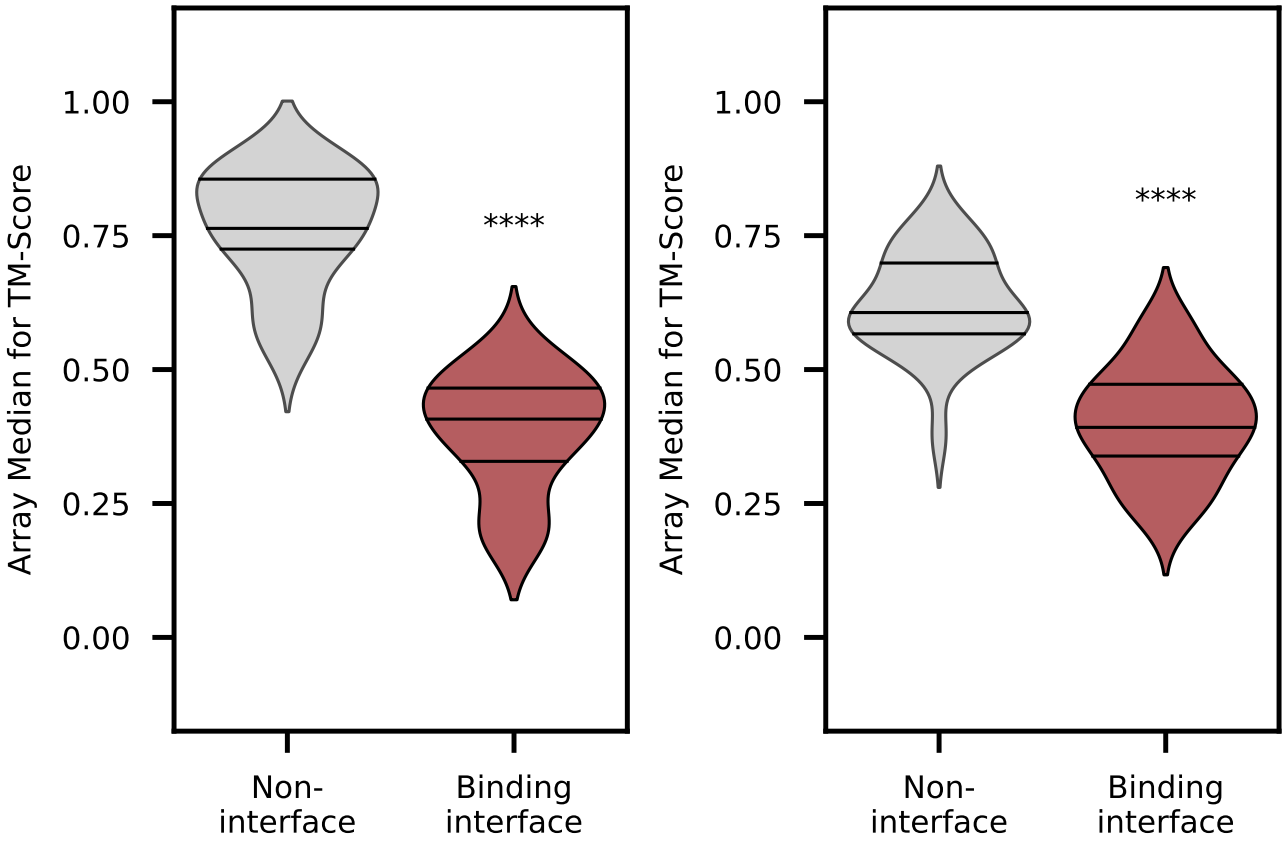

Figure 5.14: Evasin binding interfaces show lower structural similarity than non-interface regions. Left, class A evasins. Right, class B evasins. For each evasin with at least two modelled chemokine partners, pairwise TM-score matrices were calculated separately for non-interface residues and chemokine-binding interface residues, then summarised by the median pairwise score. TM-score values range from 0 to 1, with higher values indicating greater structural similarity and lower conformational variability. Violin plots show the distribution of evasin-level medians, with internal lines marking quartiles. Binding interfaces had lower median TM-scores than non-interface regions in both evasin classes, consistent with increased conformational variability at the chemokine-binding surface. ****, $p < 0.0001$ by paired Wilcoxon signed-rank test.

5.2.8 Conservation of interface residue identity

To test for differential positioning, Sørensen-Dice coefficients for interface residue sets across chemokine-evasin pairs were computed for each evasin[158]. This coefficient measures the overlap between two sets of interface residues.

Median Sørensen–Dice coefficients were 0.710 for class A evasins and 0.671 for class B evasins. These values indicate substantial but incomplete overlap between interface residue sets, with modestly greater overlap in class A. The coefficient is a set-overlap measure and is not interpreted as the percentage of residues shared. The pairwise coefficients and class-level distributions are shown in **Figure 5.15**.

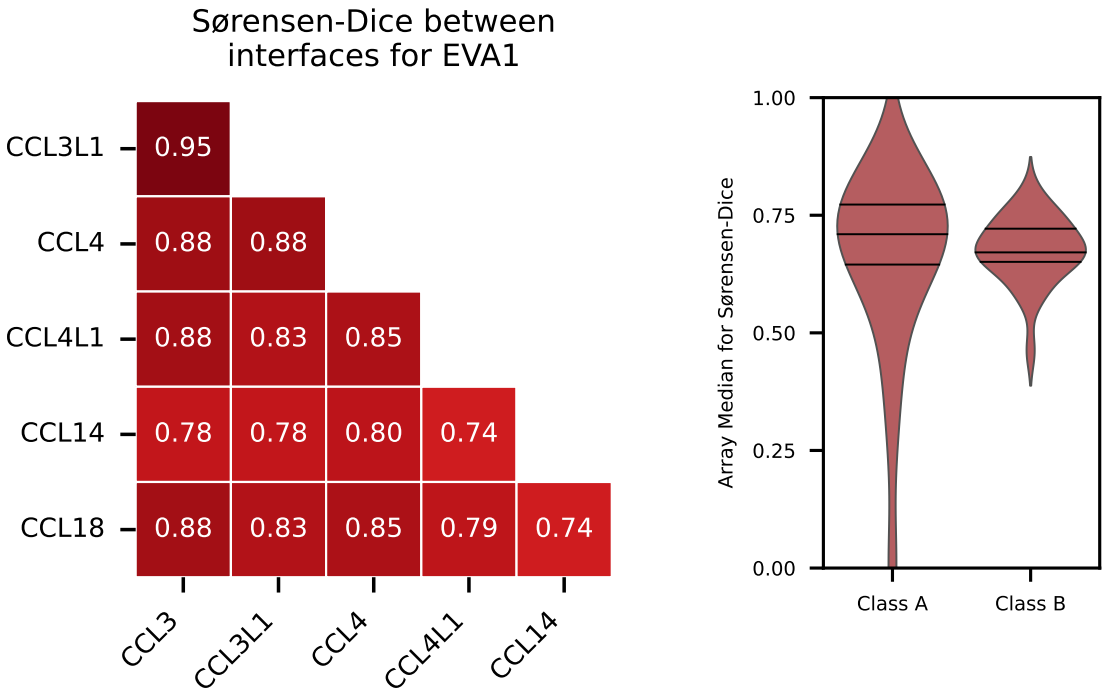

Figure 5.15: Conservation of evasin interface residue identity across chemokine partners. Left, pairwise Sørensen–Dice coefficients for EVA1 interface residue sets across six chemokine complexes. Values close to 1 indicate greater sharing of contacted evasin residues between complexes. Right, violin plots summarising the median pairwise Sørensen–Dice coefficient for each evasin with at least two modelled chemokine partners, grouped by evasin class. The distributions indicate substantial but incomplete interface-residue overlap, consistent with reuse of overlapping binding surfaces while not excluding some differential positioning.

Together, these model-derived results are consistent with conformational plasticity within substantially overlapping interfaces. The residue-set overlap does not exclude some differential positioning between complexes.

The observed coefficients indicate shared interface residues, although incomplete overlap permits partner-specific positioning. Shared residues need not contribute equally to binding, so these values alone do not establish functional equivalence between complexes. Combined with lower interface TM-scores, these results are consistent with a predicted model in which overlapping binding surfaces adopt multiple compatible shapes.

5.3 Discussion

The structural analysis of 523 chemokine interactions using AlphaFold 3 suggests how tick evasins overcome chemokine network redundancy at the molecular level. The models are consistent with broad-spectrum chemokine inhibition through conformational variability within substantially overlapping binding interfaces, while permitting some partner-specific positioning. This predicted structural mechanism is consistent with the broad chemokine binding reported for individual evasins[6].

The important point is not merely that evasins bind many chemokines, which was already known from biochemical studies[6,126]. The contribution of this analysis is to connect broad binding with a specific structural principle. A shared but flexible interface allows one compact protein scaffold to engage multiple ligands without requiring a separate rigid pocket for each. This is exactly the kind of solution expected for a redundant chemokine network: breadth is achieved through adaptable recognition of conserved chemokine features.

The quantitative analysis supports conformational plasticity as an important component of binding. Interface TM-scores are lower at binding interfaces than at non-interface regions for both evasin classes, suggesting that the binding surface is more structurally variable than the protein core. Median Sørensen–Dice coefficients

were 0.710 for class A and 0.671 for class B, indicating substantial but incomplete residue-set overlap. The models therefore favour reuse of overlapping binding surfaces but do not exclude some differential positioning.

The predicted functional consequence of evasin binding is steric occlusion of chemokine surfaces required for receptor interaction and self-association. Jaccard indices quantifying overlap between evasin-binding sites and receptor-binding sites range from 0.54 to 0.71. Permutation tests indicated that the observed overlaps were greater than expected by chance. Overlap with dimerisation sites is even greater, ranging from 0.37 to 0.88. Chemokine dimerisation has functional significance: CC chemokines cannot bind receptors as dimers because dimerisation occludes the receptor-binding surface, while CXC chemokines can bind in both forms[57,84]. Chemokine heterodimerisation can modulate receptor activation either positively or negatively[84,85]. By targeting surfaces used for both receptor interaction and self-association, evasins potentially disrupt multiple levels of chemokine signalling organisation.

This mode of inhibition differs from simply blocking one ligand-receptor pair. If an evasin binds a chemokine surface shared by receptor engagement and dimerisation, then a single binding event can prevent receptor activation, alter oligomerisation, and change the availability of the chemokine for higher-order interactions. This provides a molecular explanation for how ligand-targeted inhibitors can have broader effects than receptor-specific antagonists. It also explains why binding-site location matters: broad specificity is most useful when the inhibitor binds surfaces that are functionally central rather than peripheral.

The convergent evolution of multi-chemokine inhibitors in ticks and viruses provides evolutionary support for the network redundancy hypothesis[6,94]. The predicted conformational plasticity provides a testable hypothesis about evasin evolution. Similarly,

poxviruses, herpesviruses, and cytomegaloviruses encode secreted chemokine-binding proteins that neutralise multiple chemokines[94,95]. Tick evasins and viral chemokine-binding proteins both sequester chemokines[6,94]. This convergence suggests that broad-spectrum chemokine inhibition is a necessary strategy for organisms that must evade mammalian immune responses, implying that single-target approaches are inherently inadequate[86,95].

The contrast between evasin binding mechanisms and small-molecule chemokine receptor antagonists is instructive. Small-molecule antagonists achieve selectivity by binding deep in the receptor transmembrane pocket, exploiting sequence variation among receptors to discriminate targets[8,79]. This selectivity is a design goal for traditional drug development but becomes a liability when network redundancy enables compensation[7,8]. Evasins take the opposite approach: they bind the ligand rather than the receptor, and they prioritise breadth of ligand binding[6]. By targeting the structurally conserved features of chemokines rather than the variable features that distinguish them, evasins achieve the multi-target inhibition that network robustness demands.

Several therapeutic strategies could exploit these structural principles[86]. Evasin-derived peptides retain multi-chemokine binding activity[132,133]. Yeast-display engineering has produced antibodies that neutralise several chemokines by recognising structural features shared across the family[134]. Peptides based on the CCL5 N-terminus can disrupt chemokine heteromerisation and synergy[84]. Each strategy faces distinct challenges in terms of immunogenicity, tissue penetration, and manufacturing, but all share the conceptual foundation that multi-target inhibition is necessary to overcome network redundancy[86].

The design challenge is therefore not simply to maximise the number of chemokines

bound. A useful therapeutic inhibitor would also need to bind disease-relevant chemokines, block functionally important surfaces, avoid unwanted neutralisation of protective chemokine functions, and retain suitable pharmacokinetics. The structural principles identified here address only part of that problem: they show how breadth can be achieved and where on the chemokine surface inhibitory binding is likely to matter. Translation into a drug still requires optimisation of specificity, potency, stability, delivery, and safety.

Validated interface features could be used as constraints for de novo design of chemokine-binding proteins or peptides. Methods such as RFdiffusion provide a framework for designing new target-binding scaffolds[164]. For chemokine network inhibition, candidate scaffolds could be evaluated against several disease-relevant chemokines while avoiding binding to chemokines whose protective functions should be preserved. Whether a designed interface can accommodate several partners would require structural and functional testing.

The AlphaFold 3 analysis has limitations that should inform interpretation[146]. AlphaFold predictions are most reliable for well-folded domains and established interaction types; transient or weak interactions may be less accurately predicted[146]. The strong agreement between AlphaFold models and experimental structures in the validation set, with 41 of 42 cases showing RMSD below 2.5 Å, provides confidence that the predictions are useful for interface-level hypotheses in this protein family. AlphaFold 3 models covalent modifications, including glycosylation, but its reported benchmark excluded multi-residue glycans[146]. Glycosaminoglycan binding influences chemokine localisation and activity in vivo[57]. The analysis also cannot capture dynamic aspects of binding such as association and dissociation kinetics or conformational sampling pathways. These static models therefore generate hypotheses about interface overlap and conformational variability rather than establishing a definitive binding mechanism.

Conformational differences among the predicted complexes cannot be assigned solely to evasin flexibility. Chemokines may also undergo induced fit to optimise evasin binding, and selection among pre-existing conformations could contribute in either partner.

X-ray crystallography or solution NMR could test the predicted contact geometry, while NMR measurements or hydrogen-deuterium exchange mass spectrometry could assess changes in solution. Published multi-complex structural studies provide an experimental basis for this approach[161,162]. Mutagenesis of predicted contact residues should then be coupled to binding-affinity and kinetic measurements, receptor-signalling assays and primary-leukocyte migration assays. Testing individual chemokines and defined chemokine mixtures would establish whether structural changes affect binding breadth and functional inhibition, extending the mutagenesis-based work described previously[132,133].

The central message from this structural analysis is that evolution has produced natural inhibitors of a redundant chemokine network. Ticks and viruses independently arrived at the same functional solution: broad-spectrum binding achieved through structural mechanisms that target conserved chemokine features[6,94]. The therapeutic challenge is to translate this biological solution into clinically useful agents[86]. The predicted structural principles identified here, including conformational variability within overlapping interfaces and targeting of functional chemokine surfaces, provide a foundation for rational design of multi-chemokine inhibitors that could complement existing NF-κB and STAT-targeted therapies for immune-mediated inflammatory diseases.

Chapter 6

Discussion

6.1 Inflamed tissues show conserved immune-cell enrichment and chemokine ligand excess

Analysis of single-cell RNA-sequencing data from twelve immune-mediated inflammatory diseases showed conserved inflammatory features across distinct tissues. T-cells and macrophages constituted the majority of immune cells across the diseases, as shown in **Figure 3.4**, consistent with the central roles of adaptive and innate immune populations in chronic inflammation[11,14]. All three cytokine signalling axes showed evidence of engagement. Chemokine ligand and chemokine receptor expression was interpreted alongside immune-cell enrichment as an indirect readout of leukocyte recruitment, while fold-change analysis of pathway target genes showed changes in NF-κB targets including NFKBIA, TNFAIP3, and IL6, and STAT-associated targets including IRF1, CXCL10, and SOCS3, as shown in **Figure 3.5**[59,60,63,64]. A consistent result was the asymmetry between chemokine ligand and receptor expression. The median ligand:receptor ratio was 2.38 for CK:CKR, whereas the NF-κB and STAT medians were 0.74 and 0.55, respectively.

This result places chemokine signalling in a distinct position among the three axes analysed in this thesis. Chemokines are produced by many cell types during inflamma-

tion and act through receptors expressed on migrating leukocytes, so ligand excess is biologically compatible with their role in coordinating immune-cell traffic across tissue compartments[57,78]. The ligand excess observed here also fits the broader description of the chemokine system as a large, inducible, and overlapping signalling network rather than a set of isolated ligand-receptor pairs[58,61]. For the therapeutic question addressed in this thesis, this matters because a tissue containing many expressed chemokine ligands is unlikely to behave like a pathway in which one ligand alone controls downstream recruitment.

6.2 Chemokine networks show greater redundancy and robustness than comparator cytokine networks

Network analysis of 36 cytokine-receptor interaction networks, comprising 12 diseases multiplied by 3 signalling mechanisms, provided the central quantitative result of this thesis. Chemokine-receptor networks displayed greater edge redundancy than NF-κB or STAT networks, with median redundancy indices of 0.371, 0.124, and 0.053, respectively. In Chapter 4, this redundancy was strongly correlated with ligand-receptor promiscuity, particularly receptor-side promiscuity, explaining how the biochemical property of multi-ligand binding forms network-level architecture[61,78,83]. Robustness analyses supported the functional significance of this redundancy: chemokine-receptor networks maintained receptor activation and structural integrity longer than NF-κB or STAT networks under simulated ligand removal, potentially indicating greater tolerance for perturbation[124,125]. These results were robust to variation in the expression thresholds used for network construction.

These findings are consistent with the literature on biological robustness and network medicine. Robustness is not simply complexity. It describes the capacity of a system to preserve a specified function under perturbation, and it can arise from redundant components, alternative pathways, modular organisation, and feedback architecture[3,4]. Chemokine-driven leukocyte recruitment has previously been proposed to show robustness because many chemokines and receptors share overlapping recruitment functions, and because clinical inhibition of individual chemokine targets has produced disappointing results across multiple targets[7,8,87]. The analysis here adds a direct comparative test in human disease tissues by showing that CK:CKR networks are more redundant and robust than two cytokine axes with established therapeutic tractability.

The spatial scRNA-seq analysis adds tissue-position evidence to this comparison. In psoriasis lesional skin and ulcerative colitis inflamed rectal tissue, CK:CKR networks remained more redundant and robust than NF-κB and STAT networks when ligand-receptor edges were restricted using measured cell positions. This result links the graph analysis to the spatial organisation of inflamed tissue, which is central to chemokine-guided leukocyte recruitment.

6.3 Tick evasins reveal a structural route to broad chemokine inhibition

Structural analysis of 523 chemokine interactions using AlphaFold 3 investigated how tick evasins achieve the broad-spectrum binding necessary to overcome network redundancy. Significantly lower TM-scores at interface regions than at non-interface regions indicate greater predicted conformational variability at binding surfaces. Median Sørensen–Dice coefficients of 0.710 for class A and 0.671 for class B indicated substan-

tial but incomplete overlap between interface residue sets. Together, these results are consistent with conformational plasticity within overlapping binding surfaces while not excluding some differential positioning. Evasin binding sterically occludes chemokine surfaces required for receptor interaction and self-association, with Jaccard indices indicating significant overlap between evasin-binding sites and these functional regions.

This structural result connects the network analysis to a naturally evolved mode of chemokine inhibition. Ticks require prolonged attachment to mammalian hosts and secrete chemokine-binding proteins that reduce inflammatory recruitment at the bite site[6,126]. Evasins do not behave like highly selective receptor antagonists. They can bind chemokine ligands with multi-chemokine specificity, and several evasin-derived strategies have been proposed as routes to therapeutic chemokine network inhibition[6,86,103]. The structural analysis in this thesis suggests one mechanism by which this breadth can be achieved: substantially overlapping binding surfaces that can adopt different conformations around different chemokine partners.

6.4 Expression, topology, and structure converge on multi-target inhibition

The three results chapters form a single argument across scales of organisation. Chapter 3 establishes the transcriptional substrate: immune-mediated inflammatory diseases contain recurring immune cell populations, all three cytokine axes are detectably expressed, and the chemokine axis shows a conserved excess of expressed ligands over expressed receptors. Chapter 4 converts that substrate into disease-specific ligand-receptor graphs and shows that the CK:CKR axis has higher redundancy and robustness than the NF-κB and STAT axes. Chapter 5 then identifies how a naturally evolved inhibitor class can bind multiple chemokines through a shared but flexible

interface.

This progression matters because the chapters answer linked parts of the same therapeutic question. The scRNA-seq analysis shows expression patterns, the network analysis quantifies robustness, and the structural analysis explains how broad inhibition can overcome that robustness. The structural result has therapeutic significance because the preceding chapters establish the target network as redundant. Taken together, the chapters connect human disease expression, network architecture, and inhibitory protein structure into one mechanistic framework.

The thesis also links two different scales of redundancy. At the network scale, redundancy appears as excess ligand-receptor connectivity and persistence of receptor connectivity under simulated ligand loss. At the molecular scale, evasins respond to this redundancy by binding multiple chemokines through a shared but flexible interface. These are not the same phenomenon, but they are complementary. A redundant network creates the need for multi-target inhibition, and conformationally adaptable proteins provide one possible structural solution to that need[6,86].

The comparison with NF-κB and STAT signalling is central to this argument. Chemokine-targeted therapies have repeatedly failed in inflammatory indications, while therapies directed at TNF, IL-6 receptor, IL-17, and JAK-STAT signalling have achieved clinical use across several immune-mediated inflammatory diseases[7–11,45]. By analysing all three axes under the same disease and graph framework, this thesis places chemokine failure in a comparative context. The result is not simply that chemokine networks are complex, but that they are more redundant and robust than two cytokine axes with established therapeutic tractability.

6.5 Therapeutic implications

The network analysis provides a mechanistic framework for understanding both the failures and successes of cytokine-targeted therapies in immune-mediated inflammatory disease. The elevated redundancy and robustness of chemokine-receptor networks may explain why single-target antagonists have failed. Maraviroc blocks CCR5 in the context of HIV infection, where the therapeutic mechanism depends on preventing viral entry through a host entry co-receptor rather than disrupting a redundant inflammatory ligand-receptor network[8,107]. By contrast, when a CCR1 antagonist blocks CCR1 in rheumatoid arthritis, other receptors that bind the same or overlapping sets of chemokines may be able to maintain leukocyte recruitment[45,61,78]. The network provides backup pathways that single-target blockade may not eliminate[86,87].

The relative success of NF-κB and STAT pathway inhibitors has two complementary explanations. First, the network analysis indicates that these axes have lower intrinsic redundancy than CK:CKR networks, so their disease-specific ligand-receptor graphs contain fewer alternative routes for preserving receptor connectivity after ligand removal. This topology may make them more vulnerable to single-target or pathway-directed intervention. Second, NF-κB and STAT pathways lie upstream of chemokine production: these transcription factors directly induce chemokine gene expression in response to inflammatory stimuli[63,64]. By blocking NF-κB or STAT signalling, therapies may achieve indirect multi-chemokine inhibition through a single molecular intervention. The clinical efficacy of TNF inhibitors in reducing synovial inflammation and leukocyte infiltration could reflect this dual mechanism: direct suppression of TNF-dependent inflammatory programmes combined with indirect suppression of chemokine-driven recruitment[11,28].

The structural analysis of evasins identifies predicted design principles for therapeutics

that could target the chemokine system directly. Three features distinguish evasin binding from conventional single-target approaches. First, evasins target chemokine ligands rather than receptors; their class selectivity suggests recognition of conserved CC or CXC motifs[6]. Second, the models are consistent with conformational variability within substantially overlapping binding surfaces, allowing one evasin to accommodate multiple partners. Third, evasins are predicted to target functional chemokine surfaces required for receptor binding and self-association, potentially disrupting multiple aspects of chemokine signalling organisation.

Several therapeutic strategies incorporate these principles[86]. Evasin-derived peptides retain multi-chemokine binding activity and can be engineered as IgG-Fc fusions for improved pharmacokinetics[132,133]. Broadly neutralising antibodies engineered through yeast surface display exploit structural conservation among chemokines to achieve multi-target binding[134]. Peptides derived from the CCL5 N-terminus provide another strategy by blocking chemokine heteromerisation and its synergistic effects[84]. Each approach faces distinct development challenges, but all share the conceptual foundation that network redundancy demands multi-target intervention[86].

The therapeutic strategy implied by these findings is not to replace existing NF-κB and STAT-targeted therapies but to complement them. Many patients fail to respond adequately to TNF inhibitors, IL-6 receptor antagonists, or JAK inhibitors, while others lose response over time or develop intolerable adverse effects[10,19]. Multi-chemokine inhibitors could provide an alternative or adjunctive approach for these patients, directly targeting the leukocyte recruitment that drives tissue damage[86]. Combination therapy with NF-κB or STAT inhibitors might attack inflammation at multiple network levels, although combined cytokine suppression would require careful safety testing because existing anti-cytokine therapies already carry infection and malignancy risks[10,11].

The clinical translation of this idea requires careful target selection. A broad chemokine inhibitor is attractive only if its breadth covers disease-relevant chemokines while avoiding unacceptable suppression of protective immune surveillance. Systemic shutdown of chemokine signalling would be undesirable because chemokines contribute to host defence, immune-cell trafficking, and tissue repair[57,58,65]. Targeting anti-chemokine therapeutics to inflamed tissues could reduce this risk by concentrating inhibition where pathogenic recruitment is occurring. At the same time, the clinical use of TNF inhibitors, IL-6 receptor antagonists, IL-17 inhibitors, and JAK inhibitors indicates that cytokine signalling can be disrupted therapeutically without causing complete immune-system shutdown, provided that target selection, dosing, and patient monitoring are appropriate[9–11]. The network framework developed here could support target selection by identifying chemokines that contribute many edges across disease networks or that maintain receptor connectivity after other ligands are removed. Such prioritisation would differ from traditional single-target selection based on one highly expressed ligand or receptor. In a redundant system, the most abundant target may not be the most structurally important target.

The same logic applies to patient selection. Diseases with high chemokine network redundancy may be less likely to respond to single chemokine or receptor blockade and more likely to require multi-target inhibition[7,86,87]. Conversely, a disease state or tissue compartment with lower redundancy might remain susceptible to narrower intervention. The present thesis does not define clinical response biomarkers, but it provides a framework for developing them: patient or tissue samples could be profiled, disease-specific chemokine networks could be constructed, and redundancy or robustness metrics could be related to treatment response in future studies. This approach is consistent with the broader need to stratify patients because primary non-response and secondary loss of response remain common for existing cytokine-directed

therapies[10,19].

One practical implication is that drug development programmes should distinguish target expression from target control. A chemokine receptor may be expressed in diseased tissue and may even be biologically relevant, but it may still be a poor single target if other receptors or ligands can maintain recruitment[61,78]. Similarly, a highly expressed chemokine may be an attractive biomarker of inflammation without being a sufficient point of therapeutic control. Network position, redundancy, and robustness should therefore complement expression level when selecting targets[3,118].

Another implication concerns combination therapy. Existing anti-cytokine and JAK-inhibitor therapies act upstream or broadly within inflammatory signalling, whereas a multi-chemokine inhibitor would act more directly on leukocyte recruitment[10,11,86]. Combining these approaches could be rational in diseases where both cytokine production and cellular infiltration remain active. However, combination therapy also increases the risk of immunosuppression[10,11]. The network framework can help define which combinations are mechanistically justified, but safety and efficacy would need to be tested experimentally.

6.6 Reframing the redundancy debate

The redundancy debate has divided the chemokine field for over a decade[87,88]. Proponents of redundancy pointed to the biochemical promiscuity of chemokine-receptor interactions, the convergent evolution of multi-chemokine inhibitors in parasites and viruses, and the clinical failure of single-target antagonists[86,87,95]. Opponents emphasised context-specificity, citing mouse knockout phenotypes that revealed non-overlapping functions for chemokines and receptors that appeared redundant in vitro[88,93,112]. The available evidence therefore supported competing interpretations,

and the debate proved difficult to resolve using existing experimental approaches[113].

This analysis provides new evidence that directly addresses this controversy. Previous work had not directly tested, in human disease tissues, whether chemokine signalling networks are more redundant than comparator cytokine axes analysed under the same graph framework. By analysing cytokine expression in human disease tissues and constructing networks based on expression patterns, I demonstrate that chemokine-receptor networks in human disease display redundancy that significantly exceeds that of NF-κB and STAT networks.

The discrepancy between these findings and mouse knockout studies[88,112] can be reconciled by recognising the biological differences between these experimental contexts[114,115]. Inbred mice maintained under pathogen-free conditions represent a highly artificial situation: near-identical inherited genetic backgrounds, limited microbial exposure, naive-skewed T-cell compartments[114,115]. These conditions may minimise the functional overlap that exists under natural conditions[87]. A redundant backup system that evolved to handle the unpredictable challenges of the natural world may appear non-redundant when tested in the controlled uniformity of the laboratory[115].

Humans, in contrast, might exhibit conditions under which chemokine network redundancy is more relevant: numerous genetic variants in chemokine and receptor genes, diverse immune experience accumulated over decades of microbial exposure, and heterogeneous immune cell populations reflecting this history[115–117]. Under these conditions, the capacity for one chemokine pathway to compensate for another may be realised in ways that never occur in inbred, pathogen-free mice[87].

These findings do not invalidate the mouse knockout studies that supported context-specificity[88,112]. Individual chemokines and receptors may have non-redundant functions in specific developmental or homeostatic contexts while contributing to redundant

inflammatory networks in disease. To show definitively that context-specificity removes redundancy across the human diseases addressed in this thesis, one would need to show context-specific, non-compensatory function for every apparent alternative pathway in each relevant disease tissue. By contrast, a system only needs at least two pathways that can compensate for one another to contain redundancy. The challenge for therapeutic development is that even limited redundant signalling may be enough to preserve leukocyte recruitment after single-target inhibition[7,86].

Here, I argue that the redundancy debate should be reframed. The question is not whether chemokines are redundant or specific in an absolute sense. Biological systems can show both properties depending on scale, cell type, tissue location, and disease stage. A chemokine may have a specialised role in one context while also participating in a redundant inflammatory network in another. The more useful question for therapy is whether the disease-relevant network contains enough alternative paths to maintain function after a single target is blocked. The analyses in this thesis address that therapeutic question directly.

6.7 Evolutionary convergence

The evolutionary history of cytokine signalling helps explain why the chemokine system differs from NF-κB and STAT networks[80]. Chemokine signalling arose roughly 500 million years ago in early vertebrates[80]. Its origin postdates both Rel/NF-κB proteins, which trace to opisthokonts, and the JAK–STAT pathway, which predates chordates[159,160]. Later divergence between these systems shaped their present network properties[80].

Chemokine receptors diversified early through gene duplication events in the vertebrate stem group, establishing the major receptor families before the radiation of jawed

vertebrates[80]. Since then, receptor diversity has remained relatively stable: mammals, birds, reptiles, and fish share similar numbers of chemokine receptor genes[80]. In contrast, chemokine ligands have undergone successive expansions[80]. The first expansion accompanied the emergence of jawed vertebrates, establishing the CC and CXC subfamilies[80]. A second expansion occurred in bony fish. The most dramatic expansion occurred in placental mammals, generating paralogs such as CXCL1–3, CXCL5/6, CCL3, CCL18, and CCL25 that have no orthologs outside mammals[80]. This asymmetric expansion created the ligand-receptor imbalance that underlies biochemical promiscuity and network redundancy[61,80].

The evolutionary forces that drove chemokine ligand expansion remain incompletely understood, but host-pathogen coevolution provides a compelling hypothesis[80,94,95]. The important point is not that ticks are unique in evolving ways to shut down chemokine signalling. Multiple parasites and pathogens have independently evolved proteins that bind or disrupt chemokines, suggesting that chemokine neutralisation is a broader immune-evasion strategy[86,95]. Mammalian viruses encode secreted chemokine-binding proteins that modulate host immunity, including proteins from poxviruses, herpesviruses, and cytomegaloviruses[94]. Helminths also provide an example: *Schistosoma mansoni* secretes SmCKBP, a chemokine-binding protein that binds and blocks several chemokines including CXCL8, CCL2, CCL3, CCL5, and CX3CL1[109,165]. These examples show that broad chemokine neutralisation has evolved in ticks, viruses, and helminths, consistent with selection for suppression of chemokine-driven recruitment rather than a single isolated chemokine.

If pathogen pressure selected for chemokine diversification as a countermeasure, then the expansion of the ligand repertoire may represent an adaptive response that makes it harder for any single pathogen protein to disable the entire system[80,94]. This evolutionary dynamic would naturally generate the redundant network observed in

human disease. It would also explain the evolution of pathogen and parasite inhibitors that target multiple chemokines: a narrow inhibitor would leave alternative recruitment signals intact[86,95].

The analyses in Chapter 4 show that the chemokine system has greater ligand-receptor redundancy than the NF-κB and STAT axes. The independent evolution of multi-chemokine inhibitors in ticks, viruses, and helminths provides a natural experiment testing the redundancy hypothesis[6,94,109,165]. The predicted conformational variability provides a testable hypothesis about evasin evolution. Viruses and helminths, facing related challenges in evading mammalian immunity, converged on chemokine neutralisation through different protein folds and secretion contexts[94,109,165]. This convergence is consistent with, and perhaps supports, the view that single-target inhibition is inherently inadequate against a redundant network, and is consistent with the failures of therapeutic drugs to target chemokine signalling in inflammatory disease[7,8,86,95].

The evolutionary perspective also explains why chemokine signalling is difficult to drug using conventional specificity-driven pharmacology. Host defence systems are shaped by pressure to remain functional despite pathogen interference, tissue damage, and genetic variation. A signalling system under such pressure is expected to favour robustness and backup connectivity[3,4]. Conventional drug development, by contrast, has favoured highly selective ligands that minimise off-target effects. The mismatch between an evolved redundant target and a highly selective inhibitor may therefore be one reason chemokine pharmacology has repeatedly disappointed in inflammatory disease[7,8,45].

6.8 Systemic interpretation

The main conceptual contribution of this thesis is to treat chemokine therapeutic failure as a systemic problem rather than as a sequence of unrelated target failures. Many previous explanations for failed chemokine antagonists focus on individual drugs, individual receptors, or individual diseases[7,8,45]. Those explanations may be correct in particular cases, but they do not explain why so many different chemokine targets have failed across distinct inflammatory diseases. A network-level explanation is better matched to the pattern of failure because it asks whether the target class itself has architectural properties that resist narrow intervention[3,118].

This interpretation is grounded in a broader shift from single-component explanations of disease to network-based explanations. Network medicine argues that disease phenotypes can arise from perturbation of connected modules rather than isolated molecular defects[118]. Biological robustness theory similarly predicts that complex systems can preserve core functions through alternative routes, buffering mechanisms, and distributed control[3,4]. The chemokine system fits this conceptual frame because leukocyte recruitment is controlled by many ligands, receptors, atypical receptors, chemokine-chemokine interactions, glycosaminoglycan binding, and proteolytic processing[57,61,78]. In that setting, therapeutic failure is not surprising if the intervention removes only one component while leaving the recruitment module largely connected.

The thesis also distinguishes three related but separate ideas: expression, connectivity, and robustness. Expression shows that a ligand or receptor is present in disease tissue. Connectivity shows that expressed ligands and receptors can form curated activating relationships. Robustness shows how that network behaves when components are removed. These concepts are discussed together in cytokine biology, but they are not interchangeable. A highly expressed target is not necessarily a point of network

control, and a connected network is not necessarily robust. By moving through these levels explicitly, the thesis avoids treating expression alone as evidence of therapeutic vulnerability[86,118].

Another contribution is the direct comparison of chemokine signalling with NF-κB and STAT-linked cytokine signalling. Without comparator axes, it would be difficult to decide whether the chemokine system is unusually redundant or merely complex in the way all inflammatory systems are complex. The comparator analysis shows that the CK:CKR axis differs from two cytokine systems where single-target or pathway-directed therapies have achieved clinical success[9–11]. This comparative design makes the therapeutic interpretation stronger than a chemokine-only analysis would be.

Finally, the evasin analysis connects therapeutic strategy to a naturally evolved solution. Ticks, viruses, and helminths did not converge on selective chemokine receptor antagonism. They evolved secreted proteins that bind chemokines or otherwise interfere with chemokine-driven recruitment[6,94,109,165]. That observation is not proof of clinical efficacy, but it is a useful biological clue. The structural analysis in this thesis identifies a candidate mechanism by which broad binding can be achieved: conformational variability within substantially overlapping interfaces. This links the systemic finding of redundancy to a molecular design principle for overcoming it[86].

6.9 Limitations

Several limitations qualify the interpretation of these findings. The most fundamental is the use of RNA transcript abundance as a proxy for protein expression and functional activity. Transcript and protein abundance correlate moderately[141]. Post-transcriptional regulation, protein stability, and post-translational modification can introduce discordance. Chemokines undergo proteolytic processing that modulates

their activity[78], and receptor signalling depends on coupling efficiency that transcript levels cannot capture[57]. Proteomic validation of the networks constructed here would strengthen confidence in their biological relevance. However, I am not aware of a proteomic data resource with sufficient disease breadth, cellular resolution, cytokine measurement, and sensitivity to support the same analysis performed in this thesis. The limitation is partly technical: cytokines are low-abundance proteins that can be present at pg/mL concentrations, and mass spectrometry-based proteomics has limited sensitivity for unbiased quantification of low-abundance cytokines in complex biological samples[166]. Targeted immunoassays can measure selected cytokines with greater sensitivity; the network analysis in this thesis requires measurement of many ligands and receptors across the cytokine axes[166].

The network analyses are based on topology rather than interaction strength. I treated all expressed ligand-receptor pairs as equivalent edges, without weighting by binding affinity or signalling potency. This simplification was necessary because potency data derive from heterologous expression systems that may not reflect in vivo conditions, and comprehensive affinity measurements across all ligand-receptor pairs do not exist[61]. Incorporating quantitative interaction data could reveal whether some edges contribute more to network function than others, potentially identifying nodes whose removal would disproportionately affect network robustness.

The primary twelve-disease scRNA-seq analysis provides a snapshot of a single disease stage from dissociated datasets. Chemokine function depends critically on spatial organisation: concentration gradients formed by chemokines bound to glycosaminoglycans on endothelial surfaces and in the extracellular matrix direct leukocyte migration along defined paths[57,95]. Cells receiving the same chemokine signal may respond differently depending on their location relative to the gradient source[57]. The spatial scRNA-seq analysis in psoriasis lesional skin and ulcerative colitis inflamed rectal

tissue addressed this limitation by constructing tile-level networks from datasets that reported cell positions. The analysis provides spatial evidence from two disease contexts measured with targeted gene panels.

The dynamics of network architecture during disease progression remain unexplored. Chronic inflammatory diseases evolve over months to years, and cytokine expression patterns likely change as disease progresses from initiation through chronic inflammation to tissue remodelling[11,14]. The redundancy that characterises established disease may differ from that present during disease initiation or resolution. Longitudinal sampling would be required to address this temporal dimension.

The AlphaFold structural analysis has distinct limitations[146]. AlphaFold 3 predicts static structures and cannot capture the dynamic conformational sampling that characterises flexible binding interfaces[146]. The predictions are most reliable for well-ordered domains; intrinsically disordered regions and transient interactions may be less accurately modelled[146]. While validation against available experimental structures was encouraging, the structural conclusions should be considered hypotheses for experimental testing rather than definitive mechanistic descriptions[146].

The disease datasets also differ in sampling design, tissue processing, sequencing platform, and available controls. Applying a uniform pipeline reduces analytical variation, but it cannot make heterogeneous public datasets equivalent to a prospective multi-disease study designed from the outset. The strongest conclusions are therefore those that depend on patterns repeated across the disease set and across sensitivity analyses. Disease-specific conclusions should be treated more cautiously unless supported by the original study design and sufficient sample-level replication.

6.10 Future directions

Several research directions follow from these findings. Spatial scRNA-seq should be extended beyond the two datasets analysed here. Additional datasets with cell positions, broader measurement of cytokine and receptor genes, and more immune-mediated inflammatory diseases would test whether CK:CKR redundancy remains higher than NF-κB and STAT redundancy across more tissues. Richer spatial panels or whole-transcriptome spatial methods would also allow tile-level network analysis to include more cytokine genes and non-cytokine regulators. Such studies could identify tissue regions where network redundancy is lower and single-target approaches might succeed.

Longitudinal studies tracking cytokine network architecture through disease progression and treatment response would address the temporal dynamics that cross-sectional analysis cannot capture. Chemokine network redundancy may vary between disease initiation, established inflammation, and resolution phases. Treatment response and resistance might be associated with changes in network topology. Such studies could inform patient stratification and treatment timing[11,118].

The structural principles identified through AlphaFold modelling generate testable hypotheses for experimental structural biology. X-ray crystallography or solution NMR of the same evasin bound to several chemokine partners could test the predicted interfaces, extending existing multi-complex structural studies[161,162]. Comparisons with unbound structures and solution measurements should examine conformational changes in both the evasin and the chemokine, since induced fit by either partner remains possible. Mutagenesis should be linked to binding-affinity and kinetic measurements, receptor signalling and leukocyte migration, with controls for protein folding[132,133]. Validated contacts could subsequently guide de novo scaffold design using methods

such as RFdiffusion[164].

The translational goal is development of multi-chemokine inhibitors for clinical use[86]. Evasin-derived peptides[132,133], broadly neutralising antibodies[134], and chemokine-dimerisation inhibitors[84] are all in preclinical development. The progression of these agents into clinical trials will test whether the network redundancy principle translates into therapeutic efficacy[86]. Success would validate the conceptual framework developed here; failure would require re-evaluation of the assumptions underlying it.

Extension of the network analysis approach to additional cytokine families and disease states would provide a more complete understanding of cytokine signalling architecture in human disease. The computational framework developed here can be applied to new scRNA-seq datasets as they become available, enabling systematic comparison across conditions and patient populations.

Another important future direction is experimental perturbation of reconstructed networks. The current robustness analyses simulate ligand removal computationally. In vitro systems using primary human immune cells, tissue explants, organoids, or microfluidic migration assays could test whether predicted robust networks resist single-ligand or single-receptor blockade. Multi-chemokine inhibitors could then be compared with single-target inhibitors in the same experimental system. This would directly connect graph robustness to cellular behaviour.

A final future direction is prospective clinical-data integration. If pretreatment tissue biopsies from patients receiving cytokine-targeted therapies were profiled by single-cell or spatial transcriptomics, the network metrics developed here could be tested as predictors of response. For example, high CK:CKR redundancy might predict poor response to narrow chemokine blockade, while lower redundancy or dominance of a small ligand-receptor module might identify settings where narrower targeting remains

plausible. This would move the framework from retrospective mechanistic analysis toward clinically useful stratification.

6.11 Concluding remarks

This thesis provides evidence that chemokine-receptor networks in human immune-mediated inflammatory diseases possess redundancy and robustness that exceeds that of NF-κB and STAT networks. This network property supports a mechanistic explanation for the consistent clinical failure of single-target chemokine inhibitors despite three decades of drug development effort[7,8]. The lower redundancy of NF-κB and STAT networks, combined with their upstream position as drivers of chemokine production, may help explain why therapies targeting these pathways have achieved clinical success[9–11].

Spatially constrained graphs preserved the ordering seen in the non-spatial analysis for psoriasis lesional skin and ulcerative colitis inflamed rectal tissue. Chemokine networks exceeded both comparator axes in redundancy and robustness after ligand-receptor edges were restricted using measured cell positions.

The structural analysis of tick evasins suggests how natural selection solved the problem of attacking a redundant chemokine network. The models are consistent with broad-spectrum inhibition through conformational variability within substantially overlapping binding surfaces, while permitting some partner-specific positioning. By targeting structurally conserved chemokine surfaces required for receptor binding and self-association, evasins overcome the backup mechanisms that defeat single-target approaches[6]. The convergent evolution of related solutions in ticks, viruses, and helminths is consistent with the view that multi-target inhibition may be necessary for effective intervention against redundant chemokine networks[6,94,95,109,165].

These findings have direct implications for therapeutic development[86]. Multi-chemokine inhibitors that mimic the structural plasticity of evasins may succeed where single-target antagonists have failed. Such agents would complement rather than replace existing NF-κB and STAT-targeted therapies, addressing the substantial unmet need in patients who fail to respond to current treatments or who develop resistance over time[10,19]. The principles identified here provide a foundation for rational design of therapeutics that attack the chemokine network at its most vulnerable points[86].

The broader conclusion is that therapeutic targets should be evaluated by their biological importance and by their position within disease networks. Chemokines are clearly important mediators of leukocyte recruitment, but importance alone has not made them tractable single targets. The analyses in this thesis suggest that tractability depends on whether a target controls a pathway bottleneck or sits within a redundant network with many alternative routes. This distinction helps reconcile the central paradox of chemokine pharmacology: the system is biologically essential in inflammation, yet narrow blockade has repeatedly failed. A successful therapeutic strategy must therefore match the architecture of the target system.

References

[1] Klaus Strebhardt and Axel Ullrich. "Paul Ehrlich's magic bullet concept: 100 years of progress". In: *Nature Reviews Cancer* 8.6 (2008), pp. 473–480. DOI: 10.1038/nrc2394.

[2] Andrew L. Hopkins. "Network pharmacology: the next paradigm in drug discovery". In: *Nature Chemical Biology* 4.11 (2008), pp. 682–690. DOI: 10.1038/nchembio.118.

[3] Hiroaki Kitano. "Biological robustness". In: *Nature Reviews Genetics* 5.11 (2004), pp. 826–837. DOI: 10.1038/nrg1471.

[4] Jörg Stelling, Uwe Sauer, Zoltan Szallasi, Francis J. Doyle, and John Doyle. "Robustness of cellular functions". In: *Cell* 118.6 (2004), pp. 675–685. DOI: 10.1016/j.cell.2004.09.008.

[5] Bryan L. Roth, Douglas J. Sheffler, and Wesley K. Kroeze. "Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia". In: *Nature Reviews Drug Discovery* 3.4 (2004), pp. 353–359. DOI: 10.1038/nrd1346.

[6] Ram Prasad Bhusal, James R. O. Eaton, Sayeeda T. Chowdhury, Christine A. Power, Amanda E. I. Proudfoot, Martin J. Stone, and Shoumo Bhattacharya. "Evasins: Tick Salivary Proteins that Inhibit Mammalian Chemokines". In: *Trends in Biochemical Sciences* 45 (2020), pp. 108–122. DOI: 10.1016/j.tibs.2019.10.003.

[7] Thomas J. Schall and Amanda E. I. Proudfoot. “Overcoming hurdles in developing successful drugs targeting chemokine receptors”. In: *Nature Reviews Immunology* 11.5 (2011), pp. 355–363. DOI: 10.1038/nri2972.

[8] Roberto Solari, James E. Pease, and Malcolm Begg. “Chemokine receptors as therapeutic targets: why aren’t there more drugs?” In: *European Journal of Pharmacology* 746 (2015), pp. 363–367. DOI: 10.1016/j.ejphar.2014.06.060.

[9] Manfred Kopf, Martin F. Bachmann, and Benjamin J. Marsland. “Averting inflammation by targeting the cytokine environment”. In: *Nature Reviews Drug Discovery* 9 (2010), pp. 703–718. DOI: 10.1038/nrd2805.

[10] Josef S. Smolen, Robert B. M. Landewé, Sytske A. Bergstra, Andreas Kerschbaumer, Alexandre Sepriano, Daniel Aletaha, Roberto Caporali, Christopher J. Edwards, Kimme L. Hyrich, Janet E. Pope, et al. “EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update”. In: *Annals of the Rheumatic Diseases* 82 (2023), pp. 3–18. DOI: 10.1136/ard-2022-223356.

[11] Iain B. McInnes and Ellen M. Gravallese. “Immune-mediated inflammatory disease therapeutics: past, present and future”. In: *Nature Reviews Immunology* 21 (2021), pp. 680–686. DOI: 10.1038/s41577-021-00603-1.

[12] GBD 2019 Diseases and Injuries Collaborators. “Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019”. In: *The Lancet* 396.10258 (2020), pp. 1204–1222. DOI: 10.1016/S0140-6736(20)30925-9.

[13] GBD 2019 IMID Collaborators. “Global, regional, and national incidence of six major immune-mediated inflammatory diseases: findings from the global burden of disease study 2019”. In: *EClinicalMedicine* 64 (2023), p. 102193. DOI: 10.1016/j.eclinm.2023.102193.

[14] Mark D. Turner, Belinda Nedjai, Tara Hurst, and Daniel J. Pennington. "Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1843.11 (2014), pp. 2563–2582. DOI: 10.1016/j.bbamcr.2014.05.014.

[15] GBD 2021 Rheumatoid Arthritis Collaborators. "Global, regional, and national burden of rheumatoid arthritis, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021". In: *Lancet Rheumatology* 5 (2023), e594–e610. DOI: 10.1016/S2665-9913(23)00211-4.

[16] GBD 2017 Inflammatory Bowel Disease Collaborators. "The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017". In: *Lancet Gastroenterology & Hepatology* 5 (2020), pp. 17–30. DOI: 10.1016/S2468-1253(19)30333-4.

[17] George A. Mensah, Valentín Fuster, Christopher J. L. Murray, Gregory A. Roth, and Global Burden of Cardiovascular Diseases and Risks Collaborators. "Global Burden of Cardiovascular Diseases and Risks, 1990–2022". In: *Journal of the American College of Cardiology* 82 (2023), pp. 2350–2473. DOI: 10.1016/j.jacc.2023.11.007.

[18] Rory R. Koenen and Christian Weber. "Chemokines: established and novel targets in atherosclerosis". In: *EMBO Molecular Medicine* 3.12 (2011), pp. 713–725. DOI: 10.1002/emmm.201100183.

[19] Andrea Rubbert-Roth, Melinda Z. Szabó, Melinda Kedves, György Nagy, Fabiola Atzeni, and Piercarlo Sarzi-Puttini. "Failure of anti-TNF treatment in patients with rheumatoid arthritis: The pros and cons of the early use of alternative biological agents". In: *Autoimmunity Reviews* 18 (2019), p. 102398. DOI: 10.1016/j.autrev.2019.102398.

[20] Catherine D. Hughes, David L. Scott, Fowzia Ibrahim, and TITRATE Programme Investigators. “Intensive therapy and remissions in rheumatoid arthritis: a systematic review”. In: *BMC Musculoskeletal Disorders* 19 (2018), p. 389. DOI: 10.1186/s12891-018-2302-5.

[21] Yi Yan, Manovriti Thakur, Emiel P. C. van der Vorst, Christian Weber, and Yvonne Döring. “Targeting the chemokine network in atherosclerosis”. In: *Atherosclerosis* 330 (2021), pp. 95–106. DOI: 10.1016/j.atherosclerosis.2021.06.912.

[22] Peige Song, Zhe Fang, Hanyu Wang, Yutong Cai, Kazem Rahimi, Yajie Zhu, F. Gerald R. Fowkes, Freya J. I. Fowkes, and Igor Rudan. “Global and regional prevalence, burden, and risk factors for carotid atherosclerosis: a systematic review, meta-analysis, and modelling study”. In: *The Lancet Global Health* 8.5 (2020), e721–e729. DOI: 10.1016/S2214-109X(20)30117-0.

[23] Akira Matsumori. “Nuclear Factor-κB is a Prime Candidate for the Diagnosis and Control of Inflammatory Cardiovascular Disease”. In: *European Cardiology Review* 18 (2023), e40. DOI: 10.15420/ecr.2023.10.

[24] Girish Pathangey, Mohamed N. Allam, Mahmoud H. Abdelnabi, Dan Sorajja, Floyd David Fortuin, Kwan S. Lee, and Mayurkumar D. Bhakta. “Beyond Lipids and Platelets: A Review of Anti-Inflammatory Strategies in Secondary Prevention of Acute Coronary Syndromes”. In: *Journal of Clinical Medicine* 14.22 (2025), p. 7910. DOI: 10.3390/jcm14227910.

[25] Maria Fasolino, Gregory W. Schwartz, Abhijeet R. Patil, Aanchal Mongia, Maria L. Golson, Yue J. Wang, Ashleigh Morgan, Chengyang Liu, Jonathan Schug, Jinping Liu, et al. “Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes”. In: *Nature Metabolism* 4 (2022), pp. 284–299. DOI: 10.1038/s42255-022-00531-x.

[26] Gabriel A. Gregory, Thomas I. G. Robinson, Sarah E. Linklater, Fei Wang, Stephen Colagiuri, Carine de Beaufort, Kim C. Donaghue, International Diabetes Federation Diabetes Atlas Type 1 Diabetes in Adults Special Interest Group, Dianna J. Magliano, Jayanthi Maniam, Trevor J. Orchard, Priyanka Rai, and Graham D. Ogle. "Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study". In: *Lancet Diabetes & Endocrinology* 10 (2022), pp. 741–760. DOI: 10.1016/S2213-8587(22)00218-2.

[27] Kevan C. Herold, Thomas Delong, Ana Luisa Perdigoto, Noah Biru, Todd M. Brusko, and Lucy S. K. Walker. "The immunology of type 1 diabetes". In: *Nature Reviews Immunology* 24 (2024), pp. 435–451. DOI: 10.1038/s41577-023-00985-4.

[28] Joana Torres, Saurabh Mehandru, Jean-Frédéric Colombel, and Laurent Peyrin-Biroulet. "Crohn's disease". In: *The Lancet* 389 (2017), pp. 1741–1755. DOI: 10.1016/S0140-6736(16)31711-1.

[29] Keyvan Heydari et al. "Global prevalence and incidence of inflammatory bowel disease: a systematic review and meta-analysis of population-based studies". In: *Gastroenterology and Hepatology From Bed to Bench* 18.2 (2025), pp. 132–146. DOI: 10.22037/ghfbb.v18i2.3105.

[30] Catherine Le Berre, Sailish Honap, and Laurent Peyrin-Biroulet. "Ulcerative colitis". In: *The Lancet* 402 (2023), pp. 571–584. DOI: 10.1016/S0140-6736(23)00966-2.

[31] Remo C. Russo and Bernhard Ryffel. "The Chemokine System as a Key Regulator of Pulmonary Fibrosis: Converging Pathways in Human Idiopathic Pulmonary Fibrosis (IPF) and the Bleomycin-Induced Lung Fibrosis Model in Mice". In: *Cells* 13 (2024), p. 2058. DOI: 10.3390/cells13242058.

[32] Toby M. Maher, Elisabeth Bendstrup, Louis Dron, Jonathan Langley, Gerald Smith, Javaria Mona Khalid, Haridarshan Patel, and Michael Kreuter. "Global incidence and prevalence of idiopathic pulmonary fibrosis". In: *Respiratory Research* 22 (2021), p. 197. DOI: 10.1186/s12931-021-01791-z.

[33] Frank O. Nestle, Daniel H. Kaplan, and Jonathan Barker. "Psoriasis". In: *New England Journal of Medicine* 361 (2009), pp. 496–509. DOI: 10.1056/NEJMra0804595.

[34] Giovanni Damiani et al. "The Global, Regional, and National Burden of Psoriasis: Results and Insights From the Global Burden of Disease 2019 Study". In: *Frontiers in Medicine* 8 (2021), p. 743180. DOI: 10.3389/fmed.2021.743180.

[35] Wolf-Henning Boehncke and Michael P. Schön. "Psoriasis". In: *The Lancet* 386 (2015), pp. 983–994. DOI: 10.1016/S0140-6736(14)61909-7.

[36] European Medicines Agency. *Tremfya*. Nov. 30, 2017. URL: https://www.ema.europa.eu/en/medicines/human/EPAR/tremfya (visited on 08/19/2026).

[37] Alberta Hoi, Talia Igel, Chi Chiu Mok, and Laurent Arnaud. "Systemic lupus erythematosus". In: *The Lancet* 403 (2024), pp. 2326–2338. DOI: 10.1016/S0140-6736(24)00398-2.

[38] Jingru Tian, Dingyao Zhang, Xu Yao, Yaqing Huang, and Qianjin Lu. "Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study". In: *Annals of the Rheumatic Diseases* 82 (2023), pp. 351–356. DOI: 10.1136/ard-2022-223035.

[39] Meiling Zheng, Zhi Hu, Xiaole Mei, Lianlian Ouyang, Yang Song, Wenhui Zhou, Yi Kong, Ruifang Wu, Shijia Rao, Hai Long, et al. "Single-cell sequencing shows cellular heterogeneity of cutaneous lesions in lupus erythematosus". In: *Nature Communications* 13 (2022), p. 7489. DOI: 10.1038/s41467-022-35209-1.

[40] Peter Izmirly et al. “Population-based prevalence and incidence estimates of primary discoid lupus erythematosus from the Manhattan Lupus Surveillance Program”. In: *Lupus Science & Medicine* 6.1 (2019), e000344. DOI: 10.1136/lupus-2019-000344.

[41] Alexander J. Jafari, Christina McGee, Natasha Klimas, and Adelaide A. Hebert. “Monoclonal antibodies for the management of cutaneous lupus erythematosus: an update on the current treatment landscape”. In: *Clinical and Experimental Dermatology* 50.2 (2025), pp. 314–322. DOI: 10.1093/ced/llae374.

[42] Saba Nayar, Jason D. Turner, Saba Asam, Eanna Fennell, Matthew Pugh, Serena Colafrancesco, Onorina Berardicurti, Charlotte G. Smith, Joe Flint, Ana Teodosio, et al. “Molecular and spatial analysis of tertiary lymphoid structures in Sjögren’s syndrome”. In: *Nature Communications* 16 (2025), p. 5. DOI: 10.1038/s41467-024-54686-0.

[43] Baodong Qin, Jiaqi Wang, Zaixing Yang, Min Yang, Ning Ma, Fenglou Huang, and Renqian Zhong. “Epidemiology of primary Sjögren’s syndrome: a systematic review and meta-analysis”. In: *Annals of the Rheumatic Diseases* 74.11 (2015), pp. 1983–1989. DOI: 10.1136/annrheumdis-2014-205375.

[44] Manuel Ramos-Casals, Pilar Brito-Zerón, Stefano Bombardieri, Hendrika Bootsma, Salvatore De Vita, Thomas Dörner, Benjamin A. Fisher, Jacques-Eric Gottenberg, Gabriela Hernandez-Molina, Agnes Kocher, et al. “EULAR recommendations for the management of Sjögren’s syndrome with topical and systemic therapies”. In: *Annals of the Rheumatic Diseases* 79 (2020), pp. 3–18. DOI: 10.1136/annrheumdis-2019-216114.

[45] Zoltán Szekanecz and Alisa E. Koch. “Successes and failures of chemokine-pathway targeting in rheumatoid arthritis”. In: *Nature Reviews Rheumatology* 12 (2016), pp. 5–13. DOI: 10.1038/nrrheum.2015.157.

[46] Achilleas Floudas, Conor M. Smith, Orla Tynan, Nuno Neto, Vinod Krishna, Sarah M. Wade, Megan Hanlon, Clare Cunningham, Viviana Marzaioli, Mary Canavan, et al. “Distinct stromal and immune cell interactions shape the pathogenesis of rheumatoid and psoriatic arthritis”. In: *Annals of the Rheumatic Diseases* 81 (2022), pp. 1224–1242. DOI: 10.1136/annrheumdis-2021-221761.

[47] Lorenza Scotti, Matteo Franchi, Antonio Marchesoni, and Giovanni Corrao. “Prevalence and incidence of psoriatic arthritis: A systematic review and meta-analysis”. In: *Seminars in Arthritis and Rheumatism* 48.1 (2018), pp. 28–34. DOI: 10.1016/j.semarthrit.2018.01.003.

[48] Laure Gossec et al. “EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2023 update”. In: *Annals of the Rheumatic Diseases* 83.6 (2024), pp. 706–719. DOI: 10.1136/ard-2024-225531.

[49] Sawsan As-Sanie, Scott C. Mackenzie, Leigh Morrison, Andrew Schrepf, Krina T. Zondervan, Andrew W. Horne, and Stacey A. Missmer. “Endometriosis: A Review”. In: *JAMA* 334 (2025), pp. 64–78. DOI: 10.1001/jama.2025.2975.

[50] Marcos A. S. Fonseca, Marcela Haro, Kelly N. Wright, Xianzhi Lin, Forough Abbasi, Jennifer Sun, Lourdes Hernandez, Natasha L. Orr, Jooyoon Hong, Yunhee Choi-Kuaea, et al. “Single-cell transcriptomic analysis of endometriosis”. In: *Nature Genetics* 55 (2023), pp. 255–267. DOI: 10.1038/s41588-022-01254-1.

[51] Miguel Camba-Gómez, Laura Arosa, Oreste Gualillo, and Javier Conde-Aranda. “Chemokines and chemokine receptors in inflammatory bowel disease: Recent findings and future perspectives”. In: *Drug Discovery Today* 27 (2022), pp. 1167–1175. DOI: 10.1016/j.drudis.2021.12.004.

[52] Markus F. Neurath. “Cytokines in inflammatory bowel disease”. In: *Nature Reviews Immunology* 14.5 (2014), pp. 329–342. DOI: 10.1038/nri3661.

[53] Vaishali R. Moulton, Abel Suárez-Fueyo, Esra Meidan, Hao Li, Masayuki Mizui, and George C. Tsokos. "Pathogenesis of Human Systemic Lupus Erythematosus: A Cellular Perspective". In: *Trends in Molecular Medicine* 23 (2017), pp. 615–635. DOI: 10.1016/j.molmed.2017.05.006.

[54] Jiahe Liao, Xinbo Yu, Ziwei Huang, Qian He, Jianying Yang, Yan Zhang, Jiaqi Chen, Weijiang Song, Jing Luo, and Qingwen Tao. "Chemokines and lymphocyte homing in Sjögren's syndrome". In: *Frontiers in Immunology* 15 (2024), p. 1345381. DOI: 10.3389/fimmu.2024.1345381.

[55] Zoltán Szekanecz, Aniko Vegvari, Zoltan Szabo, and Alisa E. Koch. "Chemokines and chemokine receptors in arthritis". In: *Frontiers in Bioscience (Scholar Edition)* 2.1 (2010), pp. 153–167. DOI: 10.2741/s53.

[56] Iain B. McInnes and Georg Schett. "The pathogenesis of rheumatoid arthritis". In: *New England Journal of Medicine* 365 (2011), pp. 2205–2219. DOI: 10.1056/NEJMra1004965.

[57] Jason W. Griffith, Caroline L. Sokol, and Andrew D. Luster. "Chemokines and chemokine receptors: positioning cells for host defense and immunity". In: *Annual Review of Immunology* 32 (2014), pp. 659–702. DOI: 10.1146/annurev-immunol-032713-120145.

[58] Albert Zlotnik and Osamu Yoshie. "The chemokine superfamily revisited". In: *Immunity* 36.5 (2012), pp. 705–716. DOI: 10.1016/j.immuni.2012.05.008.

[59] Ting Liu, Lingyun Zhang, Donghyun Joo, and Shao-Cong Sun. "NF-κB signaling in inflammation". In: *Signal Transduction and Targeted Therapy* 2 (2017), p. 17023. DOI: 10.1038/sigtrans.2017.23.

[60] John J. O'Shea, Daniella M. Schwartz, Alejandro V. Villarino, Massimo Gadina, Iain B. McInnes, and Arian Laurence. "The JAK-STAT pathway: impact on

human disease and therapeutic intervention". In: *Annual Review of Medicine* 66 (2015), pp. 311–328. DOI: 10.1146/annurev-med-051113-024537.

[61] Françoise Bachelerie, Adit Ben-Baruch, Amanda M. Burkhardt, Christophe Combadiere, Joshua M. Farber, Gerard J. Graham, Richard Horuk, Alexander H. Sparre-Ulrich, Massimo Locati, Andrew D. Luster, Alberto Mantovani, Kouji Matsushima, Philip M. Murphy, Robert Nibbs, Hisayuki Nomiyama, Christine A. Power, Amanda E. I. Proudfoot, Mette M. Rosenkilde, Antal Rot, Silvano Sozzani, Marcus Thelen, Osamu Yoshie, and Albert Zlotnik. "International Union of Basic and Clinical Pharmacology. LXXXIX. Update on the extended family of chemokine receptors and introducing a new nomenclature for atypical chemokine receptors". In: *Pharmacological Reviews* 66.1 (2014), pp. 1–79. DOI: 10.1124/pr.113.007724.

[62] Ekaterini Platanitis and Thomas Decker. "Regulatory Networks Involving STATs, IRFs, and NFκB in Inflammation". In: *Frontiers in Immunology* 9 (2018), p. 2542. DOI: 10.3389/fimmu.2018.02542.

[63] Ann Richmond. "NF-κB, chemokine gene transcription and tumour growth". In: *Nature Reviews Immunology* 2 (2002), pp. 664–674. DOI: 10.1038/nri887.

[64] Simon A. Jones and Brendan J. Jenkins. "Recent insights into targeting the IL-6 cytokine family in inflammatory diseases and cancer". In: *Nature Reviews Immunology* 18 (2018), pp. 773–789. DOI: 10.1038/s41577-018-0066-7.

[65] Caroline L. Sokol and Andrew D. Luster. "The chemokine system in innate immunity". In: *Cold Spring Harbor Perspectives in Biology* 7 (2015), a016303. DOI: 10.1101/cshperspect.a016303.

[66] David Wallach. "The Tumor Necrosis Factor Family: Family Conventions and Private Idiosyncrasies". In: *Cold Spring Harbor Perspectives in Biology* 10 (2018), a028431. DOI: 10.1101/cshperspect.a028431.

[67] Charles A. Dinarello. “The IL-1 family of cytokines and receptors in rheumatic diseases”. In: *Nature Reviews Rheumatology* 15 (2019), pp. 612–632. DOI: 10.1038/s41584-019-0277-8.

[68] Mandy J. McGeachy, Daniel J. Cua, and Sarah L. Gaffen. “The IL-17 Family of Cytokines in Health and Disease”. In: *Immunity* 50 (2019), pp. 892–906. DOI: 10.1016/j.immuni.2019.03.021.

[69] Xianglan Yao and Stewart J. Levine. “Signals: Cytokines”. In: *Encyclopedia of Biological Chemistry III*. Ed. by Joseph Jez. Elsevier, 2021, pp. 48–58. DOI: 10.1016/b978-0-12-819460-7.00107-9.

[70] Warren J. Leonard, Jian-Xin Lin, and John J. O’Shea. “The γ_c Family of Cytokines: Basic Biology to Therapeutic Ramifications”. In: *Immunity* 50 (2019), pp. 832–850. DOI: 10.1016/j.immuni.2019.03.028.

[71] Michael Dougan, Glenn Dranoff, and Stephanie K. Dougan. “GM-CSF, IL-3, and IL-5 Family of Cytokines: Regulators of Inflammation”. In: *Immunity* 50 (2019), pp. 796–811. DOI: 10.1016/j.immuni.2019.03.022.

[72] Wenjun Ouyang and Anne O’Garra. “IL-10 Family Cytokines IL-10 and IL-22: from Basic Science to Clinical Translation”. In: *Immunity* 50 (2019), pp. 871–891. DOI: 10.1016/j.immuni.2019.03.020.

[73] Dario A. A. Vignali and Vijay K. Kuchroo. “IL-12 family cytokines: immunological playmakers”. In: *Nature Immunology* 13 (2012), pp. 722–728. DOI: 10.1038/ni.2366.

[74] Elia D. Tait Wojno, Christopher A. Hunter, and Jason S. Stumhofer. “The Immunobiology of the Interleukin-12 Family: Room for Discovery”. In: *Immunity* 50 (2019), pp. 851–870. DOI: 10.1016/j.immuni.2019.03.011.

[75] Helen M. Lazear, John W. Schoggins, and Michael S. Diamond. "Shared and Distinct Functions of Type I and Type III Interferons". In: *Immunity* 50 (2019), pp. 907–923. DOI: 10.1016/j.immuni.2019.03.025.

[76] Jane F. Armstrong, Elena Faccenda, Simon D. Harding, Adam J. Pawson, Christopher Southan, Joanna L. Sharman, Brice Campo, David R. Cavanagh, Stephen P. H. Alexander, Anthony P. Davenport, et al. "The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY". In: *Nucleic Acids Research* 48 (2020), pp. D1006–D1021. DOI: 10.1093/nar/gkz951.

[77] Rik Janssens, Sofie Struyf, and Paul Proost. "The unique structural and functional features of CXCL12". In: *Cellular & Molecular Immunology* 15 (2018), pp. 299–311. DOI: 10.1038/cmi.2017.107.

[78] Catherine E. Hughes and Robert J. B. Nibbs. "A guide to chemokines and their receptors". In: *The FEBS Journal* 285.16 (2018), pp. 2944–2971. DOI: 10.1111/febs.14466.

[79] Irina Kufareva, Martin Gustavsson, Yi Zheng, Bryan S. Stephens, and Tracy M. Handel. "What Do Structures Tell Us About Chemokine Receptor Function and Antagonism?" In: *Annual Review of Biophysics* 46 (2017), pp. 175–198. DOI: 10.1146/annurev-biophys-051013-022942.

[80] Alessandra Aleotti, Matthew Goulty, Clifton Lewis, Flaviano Giorgini, and Roberto Feuda. "The origin, evolution, and molecular diversity of the chemokine system". In: *Life Science Alliance* 7 (2024), e202302471. DOI: 10.26508/lsa.202302471.

[81] Jacek Lubkowski, Grzegorz Bujacz, Luis Boqué, Peter J. Domaille, Tracy M. Handel, and Alexander Wlodawer. "The structure of MCP-1 in two crystal

forms provides a rare example of variable quaternary interactions". In: *Nature Structural Biology* 4 (1997), pp. 64–69. DOI: 10.1038/nsb0197-64.

[82] G. Marius Clore, Ettore Appella, Masaki Yamada, Kouji Matsushima, and Angela M. Gronenborn. "Three-dimensional structure of interleukin 8 in solution". In: *Biochemistry* 29 (1990), pp. 1689–1696. DOI: 10.1021/bi00459a004.

[83] Andrew B. Kleist, Martyna Szpakowska, Lindsay J. Talbot, Greg Slodkowicz, Duccio Malinverni, Monica A. Thomas, Kyler S. Crawford, Daniel J. McGrail, Acacia F. Dishman, Michael J. Wedemeyer, Madison Sluter, S. Stephen Yi, Nidhi Sahni, Francis C. Peterson, Andy Chevigné, Brian F. Volkman, and M. Madan Babu. "Encoding and decoding selectivity and promiscuity in the human chemokine-GPCR interaction network". In: *Cell* 188.13 (2025), 3603–3622.e27. DOI: 10.1016/j.cell.2025.03.046.

[84] Philipp von Hundelshausen, Stijn M. Agten, Veit Eckardt, Xavier Blanchet, Martin M. Schmitt, Hans Ippel, Carlos Neideck, Kiril Bidzhekov, Julian Leberzammer, Kanin Wichapong, et al. "Chemokine interactome mapping enables tailored intervention in acute and chronic inflammation". In: *Science Translational Medicine* 9 (2017), eaah6650. DOI: 10.1126/scitranslmed.aah6650.

[85] Amanda E. I. Proudfoot and Mariagrazia Uguccioni. "Modulation of Chemokine Responses: Synergy and Cooperativity". In: *Frontiers in Immunology* 7 (2016), p. 183. DOI: 10.3389/fimmu.2016.00183.

[86] Shoumo Bhattacharya and Akane Kawamura. "Using evasins to target the chemokine network in inflammation". In: *Inflammatory Disorders, Part A*. Ed. by Rossen Donev. Vol. 119. Advances in Protein Chemistry and Structural Biology. Academic Press, 2020, pp. 1–38. DOI: 10.1016/bs.apcsb.2019.09.003.

[87] Alberto Mantovani. “Redundancy and robustness versus division of labour and specialization in innate immunity”. In: *Seminars in Immunology* 36 (2018), pp. 28–30. DOI: 10.1016/j.smim.2017.12.006.

[88] Douglas P. Dyer, Laura Medina-Ruiz, Robin Bartolini, Fabian Schuette, Catherine E. Hughes, Kenneth Pallas, Francesca Vidler, Megan K. L. Macleod, Christopher J. Kelly, Kit Ming Lee, et al. “Chemokine Receptor Redundancy and Specificity Are Context Dependent”. In: *Immunity* 50 (2019), 378–389.e5. DOI: 10.1016/j.immuni.2019.01.009.

[89] Anne Steen, Olav Larsen, Stefanie Thiele, and Mette M. Rosenkilde. “Biased and G protein-independent signaling of chemokine receptors”. In: *Frontiers in Immunology* 5 (2014), p. 277. DOI: 10.3389/fimmu.2014.00277.

[90] Jeffrey S. Smith, Lowell T. Nicholson, Jutamas Suwanpradid, Rachel A. Glenn, Nicole M. Knape, Priya Alagesan, Jaimee N. Gundry, Thomas S. Wehrman, Amber Reck Atwater, Michael D. Gunn, Amanda S. MacLeod, and Sudarshan Rajagopal. “Biased agonists of the chemokine receptor CXCR3 differentially control chemotaxis and inflammation”. In: *Science Signaling* 11.555 (2018), eaaq1075. DOI: 10.1126/scisignal.aaq1075.

[91] Amanda E. I. Proudfoot, Tracy M. Handel, Zoë Johnson, Elaine K. Lau, Patricia LiWang, Ian Clark-Lewis, Frédéric Borlat, Timothy N. C. Wells, and Marie H. Kosco-Vilbois. “Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines”. In: *Proceedings of the National Academy of Sciences* 100.4 (2003), pp. 1885–1890. DOI: 10.1073/pnas.0334864100.

[92] Antal Rot and Ulrich H. von Andrian. “Chemokines in innate and adaptive host defense: basic chemokinese grammar for immune cells”. In: *Annual Review*

of Immunology 22 (2004), pp. 891–928. DOI: 10.1146/annurev.immunol.22.012703.104543.

[93] Amanda J. L. Ridley, Yaqing Ou, Richard Karlsson, Nabina Pun, Holly L. Birchenough, Iashia Z. Mulholland, Mary L. Birch, Andrew S. MacDonald, Thomas A. Jowitt, Craig Lawless, et al. "Chemokines form complex signals during inflammation and disease that can be decoded by extracellular matrix proteoglycans". In: *Science Signaling* 16 (2023), eadf2537. DOI: 10.1126/scisignal.adf2537.

[94] Haleh Heidarieh, Bruno Hernáez, and Antonio Alcamí. "Immune modulation by virus-encoded secreted chemokine binding proteins". In: *Virus Research* 209 (2015), pp. 67–75. DOI: 10.1016/j.virusres.2015.02.028.

[95] Amanda E. I. Proudfoot, Pauline Bonvin, and Christine A. Power. "Targeting chemokines: Pathogens can, why can't we?" In: *Cytokine* 74.2 (2015), pp. 259–267. DOI: 10.1016/j.cyto.2015.02.011.

[96] Gregory D. Wiens and Gavin W. Glenney. "Origin and evolution of TNF and TNF receptor superfamilies". In: *Developmental and Comparative Immunology* 35.12 (2011), pp. 1324–1335. DOI: 10.1016/j.dci.2011.03.031.

[97] Jack Rivers-Auty, Michael J. D. Daniels, Isaac Colliver, David L. Robertson, and David Brough. "Redefining the ancestral origins of the interleukin-1 superfamily". In: *Nature Communications* 9 (2018), p. 1156. DOI: 10.1038/s41467-018-03362-1.

[98] Yvonne Döring, Emiel P. C. van der Vorst, and Christian Weber. "Targeting immune cell recruitment in atherosclerosis". In: *Nature Reviews Cardiology* 21 (2024), pp. 824–840. DOI: 10.1038/s41569-024-01023-z.

[99] Paul M. Ridker, Brendan M. Everett, Tom Thuren, Jean G. MacFadyen, William H. Chang, Christie Ballantyne, Francisco Fonseca, Jose Nicolau, Wolfgang

Koenig, Stefan D. Anker, et al. “Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease”. In: *New England Journal of Medicine* 377 (2017), pp. 1119–1131. DOI: 10.1056/NEJMoa1707914.

[100] Palak J. Trivedi and David H. Adams. “Chemokines and Chemokine Receptors as Therapeutic Targets in Inflammatory Bowel Disease: Pitfalls and Promise”. In: *Journal of Crohn's and Colitis* 12 (2018), S641–S652. DOI: 10.1093/ecco-jcc/jjx145.

[101] GlaxoSmithKline. *Study to Evaluate the Pharmacodynamics of SB-656933 in Patients With Ulcerative Colitis*. ClinicalTrials.gov identifier NCT00748410. ClinicalTrials.gov. Oct. 30, 2020. URL: https://clinicaltrials.gov/study/NCT00748410 (visited on 08/19/2026).

[102] Paul M. O'Byrne, Hristo Metev, Margareta Puu, Kai Richter, Christina Keen, Mohib Uddin, Bengt Larsson, Marie Cullberg, and Parameswaran Nair. “Efficacy and safety of a CXCR2 antagonist, AZD5069, in patients with uncontrolled persistent asthma: a randomised, double-blind, placebo-controlled trial”. In: *The Lancet Respiratory Medicine* 4.10 (2016), pp. 797–806. DOI: 10.1016/S2213-2600(16)30227-2.

[103] Pauline Bonvin, Christine A. Power, and Amanda E. I. Proudfoot. “Evasins: Therapeutic Potential of a New Family of Chemokine-Binding Proteins from Ticks”. In: *Frontiers in Immunology* 7 (2016), p. 208. DOI: 10.3389/fimmu.2016.00208.

[104] Olaf Wolkenhauer and Allan Muir. “The Complexity of Cell-Biological Systems”. In: *Philosophy of Complex Systems*. Ed. by Cliff Hooker. Elsevier, 2011, pp. 355–385. DOI: 10.1016/b978-0-444-52076-0.50013-4.

[105] Jean M. Carlson and John Doyle. “Complexity and robustness”. In: *Proceedings of the National Academy of Sciences* 99 (2002), pp. 2538–2545. DOI: 10.1073/pnas.012582499.

[106] James M. Whitacre. “Biological robustness: paradigms, mechanisms, and systems principles”. In: *Frontiers in Genetics* 3 (2012), p. 67. DOI: 10.3389/fgene.2012.00067.

[107] J. Claiborne Stephens, David E. Reich, David B. Goldstein, Hyoung Doo Shin, Michael W. Smith, Mary Carrington, Cheryl Winkler, Gavin A. Huttley, Rando Allikmets, Lynn Schriml, et al. “Dating the origin of the CCR5-Δ32 AIDS-resistance allele by the coalescence of haplotypes”. In: *American Journal of Human Genetics* 62 (1998), pp. 1507–1515. DOI: 10.1086/301867.

[108] David H. McDermott, Alan M. Fong, Qiong Yang, Joan M. Sechler, L. Adrienne Cupples, Maya N. Merrell, Peter W. F. Wilson, Ralph B. D’Agostino, Christopher J. O’Donnell, Dhavalkumar D. Patel, and Philip M. Murphy. “Chemokine receptor mutant CX3CR1-M280 has impaired adhesive function and correlates with protection from cardiovascular disease in humans”. In: *Journal of Clinical Investigation* 111 (2003), pp. 1241–1250. DOI: 10.1172/JCI16790.

[109] Rick M. Maizels, Hermelijn H. Smits, and Henry J. McSorley. “Modulation of Host Immunity by Helminths: The Expanding Repertoire of Parasite Effector Molecules”. In: *Immunity* 49 (2018), pp. 801–818. DOI: 10.1016/j.immuni.2018.10.016.

[110] Allison J. Bancroft, Colin W. Levy, Thomas A. Jowitt, Kelly S. Hayes, Seona Thompson, Edward A. McKenzie, Matthew D. Ball, Eamon Dubaissi, Aidan P. France, Bruno Bellina, et al. “The major secreted protein of the whipworm parasite tethers to matrix and inhibits interleukin-13 function”. In: *Nature Communications* 10 (2019), p. 2344. DOI: 10.1038/s41467-019-09996-z.

[111] Alessandra Ricciardi, Sergio A. Hassan, Olena Kamenyeva, Sasisekhar Bennuru, John Andersen, and Thomas B. Nutman. "A filarial parasite-encoded human IL-10 receptor antagonist reveals a novel strategy to modulate host responses". In: *PNAS Nexus* 1 (2022), pgac184. DOI: 10.1093/pnasnexus/pgac184.

[112] Mieke Metzemaekers, Vincent Vanheule, Rik Janssens, Sofie Struyf, and Paul Proost. "Overview of the Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-Inducible CXC Chemokine Receptor 3 Ligands". In: *Frontiers in Immunology* 8 (2018), p. 1970. DOI: 10.3389/fimmu.2017.01970.

[113] Douglas P. Dyer. "Understanding the mechanisms that facilitate specificity, not redundancy, of chemokine-mediated leukocyte recruitment". In: *Immunology* 160 (2020), pp. 336–344. DOI: 10.1111/imm.13200.

[114] Jessica B. Graham, Jessica L. Swarts, Sarah R. Leist, Alexandra Schäfer, Timothy A. Bell, Pablo Hock, Joe Farrington, Ginger D. Shaw, Martin T. Ferris, Fernando Pardo-Manuel de Villena, et al. "Unique immune profiles in collaborative cross mice linked to survival and viral clearance upon infection". In: *iScience* 27 (2024), p. 109103. DOI: 10.1016/j.isci.2024.109103.

[115] Sara E. Hamilton, Vladimir P. Badovinac, Lalit K. Beura, Mark Pierson, Stephen C. Jameson, David Masopust, and Thomas S. Griffith. "New Insights into the Immune System Using Dirty Mice". In: *Journal of Immunology* 205 (2020), pp. 3–11. DOI: 10.4049/jimmunol.2000171.

[116] Charles P. Gerba. "Environmentally Transmitted Pathogens". In: *Environmental Microbiology*. Ed. by I. L. Pepper, C. P. Gerba, and T. J. Gentry. Elsevier, 2015, pp. 509–550. DOI: 10.1016/b978-0-12-394626-3.00022-3.

[117] Hanli Xu, Shuye Lin, Ziyun Zhou, Duoduo Li, Xiting Zhang, Muhan Yu, Ruoyi Zhao, Yiheng Wang, Junru Qian, Xinyi Li, Bohan Li, Chuhan Wei, Keqiang

Chen, Teizo Yoshimura, Ji Ming Wang, and Jiaqiang Huang. “New genetic and epigenetic insights into the chemokine system: the latest discoveries aiding progression toward precision medicine”. In: *Cellular & Molecular Immunology* 20 (2023), pp. 739–776. DOI: 10.1038/s41423-023-01032-x.

[118] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. “Network medicine: a network-based approach to human disease”. In: *Nature Reviews Genetics* 12.1 (2011), pp. 56–68. DOI: 10.1038/nrg2918.

[119] Jordi Bascompte and Pedro Jordano. “Plant-Animal Mutualistic Networks: The Architecture of Biodiversity”. In: *Annual Review of Ecology, Evolution, and Systematics* 38 (2007), pp. 567–593. DOI: 10.1146/annurev.ecolsys.38.091206.095818.

[120] Matthias Schleuning, Jochen Fründ, Oliver Schweiger, Erik Welk, Jörg Albrecht, Matthias Albrecht, Marion Beil, Gita Benadi, Nico Blüthgen, Helge Bruelheide, et al. “Ecological networks are more sensitive to plant than to animal extinction under climate change”. In: *Nature Communications* 7 (2016), p. 13965. DOI: 10.1038/ncomms13965.

[121] Réka Albert, Hawoong Jeong, and Albert-László Barabási. “Error and attack tolerance of complex networks”. In: *Nature* 406 (2000), pp. 378–382. DOI: 10.1038/35019019.

[122] Oriol Artime, Marco Grassia, Manlio De Domenico, James P. Gleeson, Hernán A. Makse, Giuseppe Mangioni, Matjaž Perc, and Filippo Radicchi. “Robustness and resilience of complex networks”. In: *Nature Reviews Physics* 6 (2024), pp. 114–131. DOI: 10.1038/s42254-023-00676-y.

[123] Denis Cornaz and Philippe Meurdesoif. “Chromatic Gallai identities operating on Lovász number”. In: *Mathematical Programming* 144 (2013), pp. 347–368. DOI: 10.1007/s10107-013-0636-1.

[124] Jane Memmott, Nickolas M. Waser, and Mary V. Price. “Tolerance of pollination networks to species extinctions”. In: *Proceedings of the Royal Society B: Biological Sciences* 271 (2004), pp. 2605–2611. DOI: 10.1098/rspb.2004.2909.

[125] Christian M. Schneider, André A. Moreira, José S. Andrade Jr., Shlomo Havlin, and Hans J. Herrmann. “Mitigation of malicious attacks on networks”. In: *Proceedings of the National Academy of Sciences* 108 (2011), pp. 3838–3841. DOI: 10.1073/pnas.1009440108.

[126] Shoumo Bhattacharya and Patricia A. Nuttall. “Phylogenetic Analysis Indicates That Evasin-Like Proteins of Ixodid Ticks Fall Into Three Distinct Classes”. In: *Frontiers in Cellular and Infection Microbiology* 11 (2021), p. 769542. DOI: 10.3389/fcimb.2021.769542.

[127] Jenni Hayward, Julie Sanchez, Andrew Perry, Cheng Huang, Manuel Rodriguez-Valle, Meritxell Canals, Richard J. Payne, and Martin J. Stone. “Ticks from diverse genera encode chemokine-inhibitory evasin proteins”. In: *Journal of Biological Chemistry* 292.38 (2017), pp. 15670–15680. DOI: 10.1074/jbc.M117.807255.

[128] Kamayani Singh, Graham Davies, Yara Alenazi, James R. O. Eaton, Akane Kawamura, and Shoumo Bhattacharya. “Yeast surface display identifies a family of evasins from ticks with novel polyvalent CC chemokine-binding activities”. In: *Scientific Reports* 7 (2017), p. 4267. DOI: 10.1038/s41598-017-04378-1.

[129] Maud Déruaz, Achim Frauenschuh, Ana L. Alessandri, João M. Dias, Fernanda M. Coelho, Remo C. Russo, Beatriz R. Ferreira, Gerard J. Graham, Jeffrey P. Shaw, Timothy N. C. Wells, Mauro M. Teixeira, Christine A. Power, and Amanda E. I. Proudfoot. “Ticks produce highly selective chemokine binding proteins with antiinflammatory activity”. In: *Journal of Experimental Medicine* 205 (2008), pp. 2019–2031. DOI: 10.1084/jem.20072689.

[130] João M. Dias, Christophe Losberger, Maud Déruaz, Christine A. Power, Amanda E. I. Proudfoot, and Jeffrey P. Shaw. "Structural Basis of Chemokine Sequestration by a Tick Chemokine Binding Protein: The Crystal Structure of the Complex between Evasin-1 and CCL3". In: *PLoS ONE* 4.12 (2009), e8514. DOI: 10.1371/journal.pone.0008514.

[131] Angela W. Lee, Maud Déruaz, Christopher Lynch, Graham Davies, Kamayani Singh, Yara Alenazi, James R. O. Eaton, Akane Kawamura, Jeffrey Shaw, Amanda E. I. Proudfoot, João M. Dias, and Shoumo Bhattacharya. "A knottin scaffold directs the CXC-chemokine-binding specificity of tick evasins". In: *Journal of Biological Chemistry* 294 (2019), pp. 11199–11212. DOI: 10.1074/jbc.RA119.008817.

[132] Serena Vales, Jhanna Kryukova, Soumyanetra Chandra, Gintare Smagurauskaite, Megan Payne, Charlie J. Clark, Katrin Hafner, Philomena Mburu, Stepan Denisov, Graham Davies, Carlos Outeiral, Charlotte M. Deane, Garrett M. Morris, and Shoumo Bhattacharya. "Discovery and pharmacophoric characterization of chemokine network inhibitors using phage-display, saturation mutagenesis and computational modelling". In: *Nature Communications* 14 (2023), p. 5763. DOI: 10.1038/s41467-023-41488-z.

[133] Jhanna Kryukova, Serena Vales, Megan Payne, Gintare Smagurauskaite, Soumyanetra Chandra, Charlie J. Clark, Graham Davies, and Shoumo Bhattacharya. "Development of chemokine network inhibitors using combinatorial saturation mutagenesis". In: *Communications Biology* 8 (2025), p. 549. DOI: 10.1038/s42003-025-07778-6.

[134] Alessandro Angelini, Yoshishige Miyabe, Daniel Newsted, Byron H. Kwan, Chie Miyabe, Ryan L. Kelly, Misha N. Jamy, Andrew D. Luster, and K. Dane Wittrup. "Directed evolution of broadly crossreactive chemokine-blocking anti-

bodies efficacious in arthritis". In: *Nature Communications* 9 (2018), p. 1461. DOI: 10.1038/s41467-018-03687-x.

[135] Efthymia Papalexi and Rahul Satija. "Single-cell RNA sequencing to explore immune cell heterogeneity". In: *Nature Reviews Immunology* 18.1 (2018), pp. 35–45. DOI: 10.1038/nri.2017.76.

[136] Suoqin Jin, Christian F. Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V. Plikus, and Qing Nie. "Inference and analysis of cell-cell communication using CellChat". In: *Nature Communications* 12 (2021), p. 1088. DOI: 10.1038/s41467-021-21246-9.

[137] Daniel Dimitrov, Dénes Türei, Martín Garrido-Rodríguez, Paul L. Burmedi, James S. Nagai, Charlotte Boys, Ricardo O. Ramirez Flores, Hyojin Kim, Bence Szalai, Ivan G. Costa, Alberto Valdeolivas, Aurélien Dugourd, and Julio Saez-Rodriguez. "Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data". In: *Nature Communications* 13 (2022), p. 3224. DOI: 10.1038/s41467-022-30755-0.

[138] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. "UMAP: Uniform Manifold Approximation and Projection". In: *Journal of Open Source Software* 3.29 (2018), p. 861. DOI: 10.21105/joss.00861. URL: https://doi.org/10.21105/joss.00861.

[139] C. Domínguez Conde, C. Xu, L. B. Jarvis, D. B. Rainbow, S. B. Wells, T. Gomes, S. K. Howlett, O. Suchanek, K. Polanski, H. W. King, et al. "Cross-tissue immune cell analysis reveals tissue-specific features in humans". In: *Science* 376.6594 (2022), eabl5197. DOI: 10.1126/science.abl5197.

[140] Stephanie C. Hicks, F. William Townes, Mingxiang Teng, and Rafael A. Irizarry. "Missing data and technical variability in single-cell RNA-sequencing

experiments". In: *Biostatistics* 19.4 (2018), pp. 562–578. DOI: 10.1093/biostatistics/kxx053.

[141] Anatole Ghazalpour, Brian Bennett, Vladislav A. Petyuk, Luz Orozco, Raffi Hagopian, Imran N. Mungrue, Charles R. Farber, Janet Sinsheimer, Hyun Min Kang, Nicholas Furlotte, et al. "Comparative analysis of proteome and transcriptome variation in mouse". In: *PLoS Genetics* 7.6 (2011), e1001393. DOI: 10.1371/journal.pgen.1001393.

[142] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

[143] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, et al. "Applying and improving AlphaFold at CASP14". In: *Proteins: Structure, Function, and Bioinformatics* 89.12 (2021), pp. 1711–1721. DOI: 10.1002/prot.26257.

[144] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. "Protein complex prediction with AlphaFold-Multimer". In: *bioRxiv* (2021). DOI: 10.1101/2021.10.04.463034.

[145] Marc F. Lensink, Guillaume Brysbaert, Nessim Raouraoua, Paul A. Bates, Marco Giulini, Rodrigo V. Honorato, Charlotte van Noort, João M. C. Teixeira, Alexandre M. J. J. Bonvin, Ren Kong, et al. "Impact of AlphaFold on structure

prediction of protein complexes: The CASP15-CAPRI experiment". In: *Proteins: Structure, Function, and Bioinformatics* 91.12 (2023), pp. 1658–1683. DOI: 10.1002/prot.26609.

[146] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, et al. "Accurate structure prediction of biomolecular interactions with AlphaFold 3". In: *Nature* 630 (2024), pp. 493–500. DOI: 10.1038/s41586-024-07487-w.

[147] Dima Kozakov, David R. Hall, Bing Xia, Kathryn A. Porter, Dzmitry Padhorny, Christine Yueh, Dmitri Beglov, and Sandor Vajda. "The ClusPro web server for protein-protein docking". In: *Nature Protocols* 12.2 (2017), pp. 255–278. DOI: 10.1038/nprot.2016.169.

[148] Scott A. Hollingsworth and Ron O. Dror. "Molecular Dynamics Simulation for All". In: *Neuron* 99.6 (2018), pp. 1129–1143. DOI: 10.1016/j.neuron.2018.08.011.

[149] Lejla Svraka, Anna Skarnvad Andersen, Toke Touborg, Thomas Emmanuel, Udayaraja GK, Haja N. Kadarmideen, Trine Bertelsen, Christian Vestergaard, and Claus Johansen. "IL-17 Cytokines Induce IκBζ in Dermal Fibroblasts to Promote Pro-Inflammatory Gene Expression in Psoriasis". In: *International Journal of Molecular Sciences* 27.3 (2026), p. 1297. DOI: 10.3390/ijms27031297.

[150] Yiming Li, Cenfu Wei, Wenjing Yang, Hai Wang, Alan Robinson, Isabella Peshek, Tianming Yu, Jiwoon Park, Jenny Yanyi Ding, Stephen B. Hanauer, et al. "Spatial transcriptomics atlas of inflammatory bowel disease to guide implementation in research consortiums and clinical trials". In: *Nature Communications* 17 (2026), p. 5808. DOI: 10.1038/s41467-026-72482-w.

[151] Tom Alsaigh, Doug Evans, David Frankel, and Ali Torkamani. “Decoding the transcriptome of calcified atherosclerotic plaque at single-cell resolution”. In: *Communications Biology* 5 (2022), p. 1084. DOI: 10.1038/s42003-022-04056-7.

[152] Alba Garrido-Trigo, Ana M. Corraliza, Marisol Veny, Isabella Dotti, Elisa Melón-Ardanaz, Aina Rill, Helena L. Crowell, Ángel Corbí, Victoria Gudiño, Miriam Esteller, et al. “Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease”. In: *Nature Communications* 14 (2023), p. 4506. DOI: 10.1038/s41467-023-40156-6.

[153] Arun C. Habermann, Austin J. Gutierrez, Linh T. Bui, Stephanie L. Yahn, Nichelle I. Winters, Carla L. Calvi, Lance Peter, Mei-I Chung, Chase J. Taylor, Christopher Jetter, Latha Raju, Jamie Roberson, Guixiao Ding, Lori Wood, Jennifer M. S. Sucre, Bradley W. Richmond, Ana P. Serezani, Wyatt J. McDonnell, Simon B. Mallal, Matthew J. Bacchetta, James E. Loyd, Ciara M. Shaver, Lorraine B. Ware, Ross Bremner, Rajat Walia, Timothy S. Blackwell, Nicholas E. Banovich, and Jonathan A. Kropski. “Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis”. In: *Science Advances* 6.28 (2020), eaba1972. DOI: 10.1126/sciadv.aba1972.

[154] Feiyang Ma, Olesya Plazyo, Allison C. Billi, Lam C. Tsoi, Xianying Xing, Rachael Wasikowski, Mehrnaz Gharaee-Kermani, Grace Hile, Yanyun Jiang, Paul W. Harms, et al. “Single cell and spatial sequencing define processes by which keratinocytes and fibroblasts amplify inflammatory responses in psoriasis”. In: *Nature Communications* 14 (2023), p. 3455. DOI: 10.1038/s41467-023-39020-4.

[155] Achim Frauenschuh, Christine A. Power, Maud Déruaz, Beatriz R. Ferreira, João S. Silva, Mauro M. Teixeira, João M. Dias, Thierry Martin, Timothy N. C. Wells, and Amanda E. I. Proudfoot. "Molecular cloning and characterization of a highly selective chemokine-binding protein from the tick *Rhipicephalus sanguineus*". In: *Journal of Biological Chemistry* 282.37 (2007), pp. 27250–27258. DOI: 10.1074/jbc.M704706200.

[156] Paul Jaccard. "The distribution of the flora in the alpine zone". In: *New Phytologist* 11 (1912), pp. 37–50. DOI: 10.1111/j.1469-8137.1912.tb05611.x.

[157] Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. "US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes". In: *Nature Methods* 19 (2022), pp. 1109–1115. DOI: 10.1038/s41592-022-01585-1.

[158] Lee R. Dice. "Measures of the Amount of Ecologic Association Between Species". In: *Ecology* 26 (1945), pp. 297–302. DOI: 10.2307/1932409.

[159] Michelle M. Leger et al. "Rel/NF-κB Transcription Factors Emerged at the Onset of Opisthokonts". In: *Genome Biology and Evolution* 14.1 (2022), evab289. DOI: 10.1093/gbe/evab289.

[160] Clifford Liongue, Lynda A. O'Sullivan, Monique C. Trengove, and Alister C. Ward. "Evolution of JAK-STAT Pathway Components: Mechanisms and Role in Immune System Development". In: *PLoS ONE* 7.3 (2012), e32777. DOI: 10.1371/journal.pone.0032777.

[161] Ram Prasad Bhusal, Pramod Aryal, Shankar Raj Devkota, Rina Pokhrel, Menachem J. Gunzburg, Simon R. Foster, Herman D. Lim, Richard J. Payne, Matthew C. J. Wilce, and Martin J. Stone. "Structure-guided engineering of tick evasins for targeting chemokines in inflammatory diseases". In: *Proceedings*

of the National Academy of Sciences 119.9 (2022), e2122105119. DOI: 10.1073/pnas.2122105119.

[162] Shankar Raj Devkota, Pramod Aryal, Rina Pokhrel, Wanting Jiao, Andrew Perry, Santosh Panjikar, Richard J. Payne, Matthew C. J. Wilce, Ram Prasad Bhusal, and Martin J. Stone. "Engineering broad-spectrum inhibitors of inflammatory chemokines from subclass A3 tick evasins". In: *Nature Communications* 14 (2023), p. 4204. DOI: 10.1038/s41467-023-39879-3.

[163] Shankar Raj Devkota, Pramod Aryal, Matthew C. J. Wilce, Richard J. Payne, Martin J. Stone, and Ram Prasad Bhusal. "Structural basis of chemokine recognition by the class A3 tick evasin EVA-ACA1001". In: *Protein Science* 33.6 (2024), e4999. DOI: 10.1002/pro.4999.

[164] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, et al. "De novo design of protein structure and function with RFdiffusion". In: *Nature* 620 (2023), pp. 1089–1100. DOI: 10.1038/s41586-023-06415-8.

[165] Philip Smith, Rosie E. Fallon, Niamh E. Mangan, Caitriona M. Walsh, Margarida Saraiva, Jon R. Sayers, Andrew N. J. McKenzie, Antonio Alcami, and Padraic G. Fallon. "*Schistosoma mansoni* secretes a chemokine binding protein with antiinflammatory activity". In: *Journal of Experimental Medicine* 202.10 (2005), pp. 1319–1325. DOI: 10.1084/jem.20050955.

[166] Julie A. Stenken and Andreas J. Poschenrieder. "Bioanalytical Chemistry of Cytokines: A Review". In: *Analytica Chimica Acta* 853 (2015), pp. 95–115. DOI: 10.1016/j.aca.2014.10.009.

Appendix A

Disease-specific cytokine networks

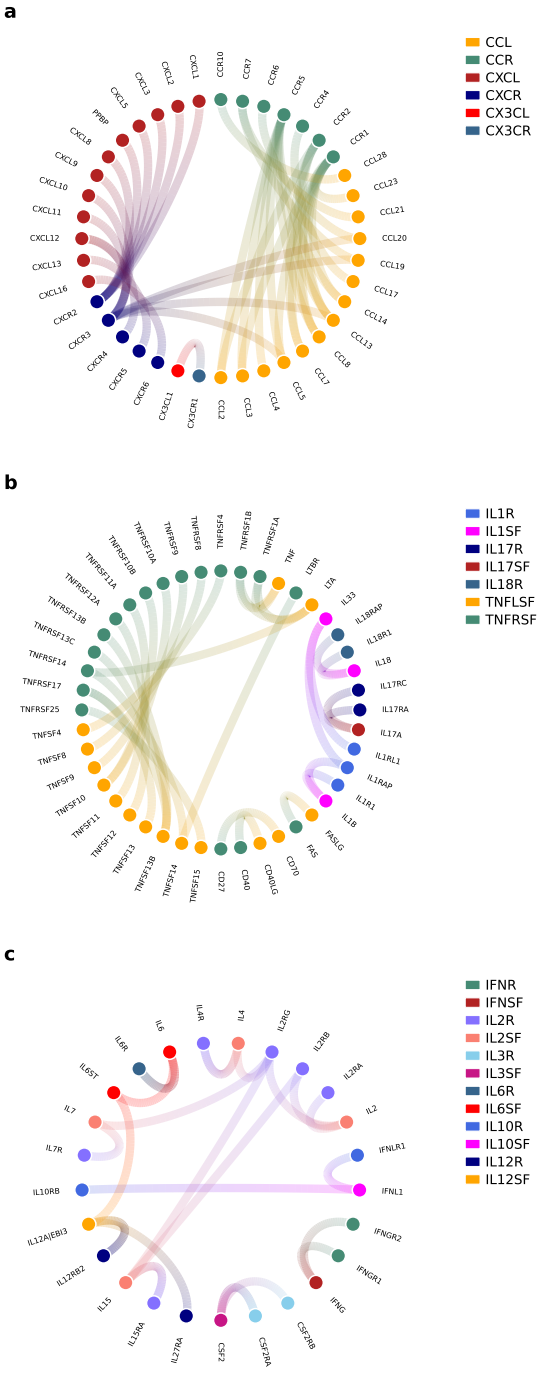

Figure A.1: Atherosclerosis in carotid tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

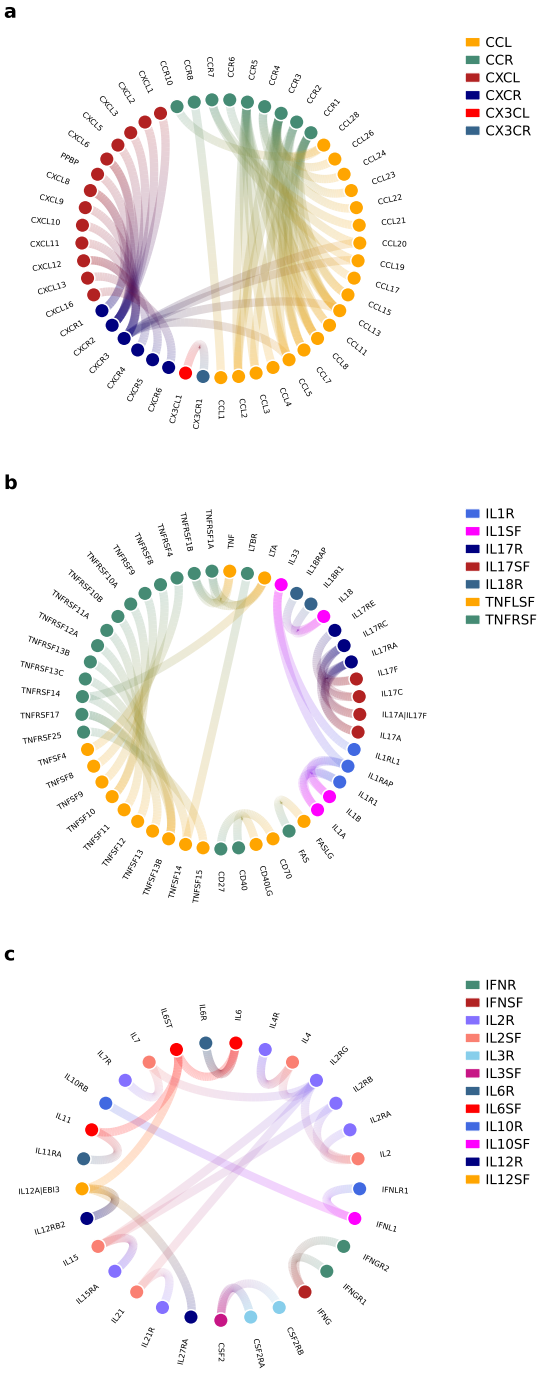

Figure A.2: Crohn's disease in colon tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

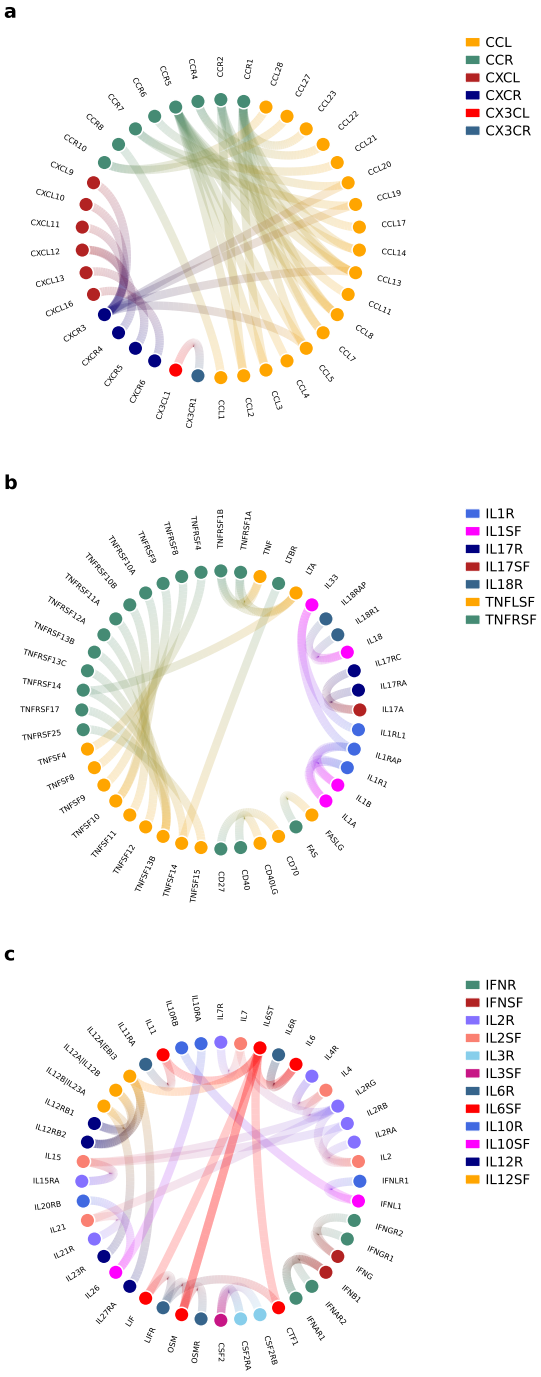

Figure A.3: Discoid lupus erythematosus in skin tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

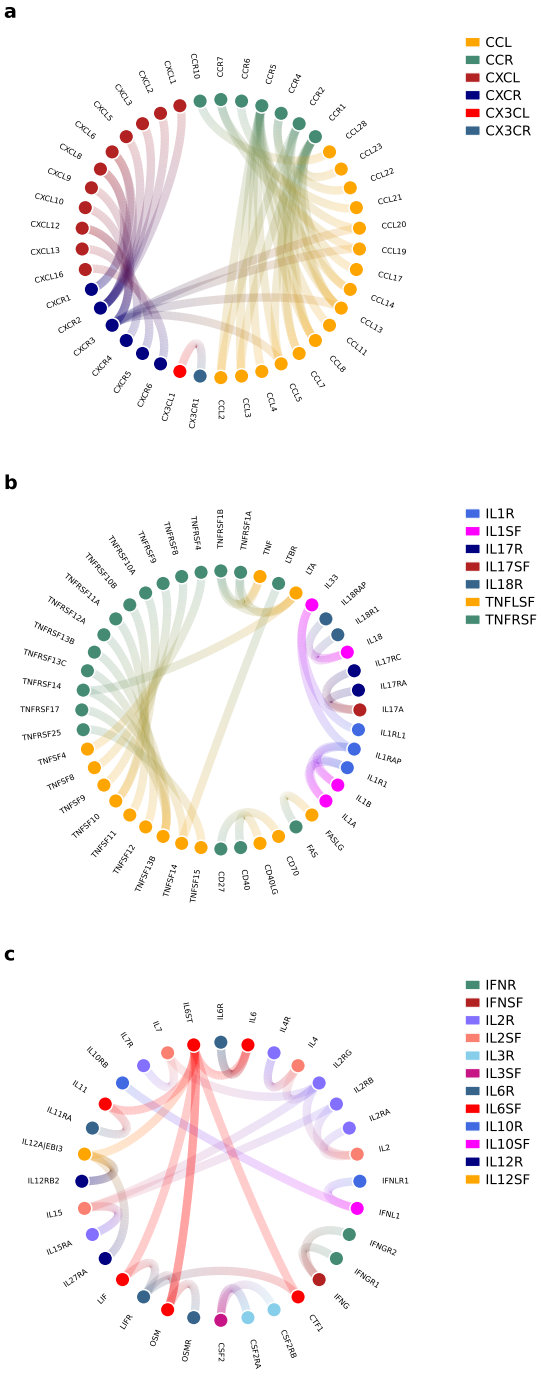

Figure A.4: Endometriosis in endometrial tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

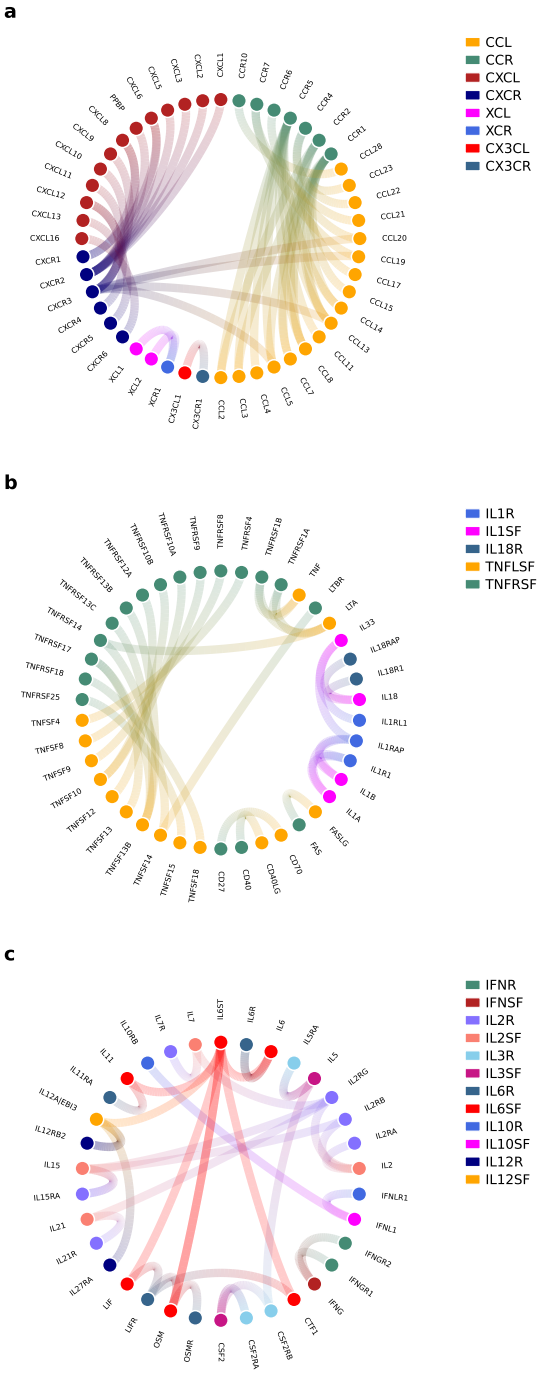

Figure A.5: Idiopathic pulmonary fibrosis in lung tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

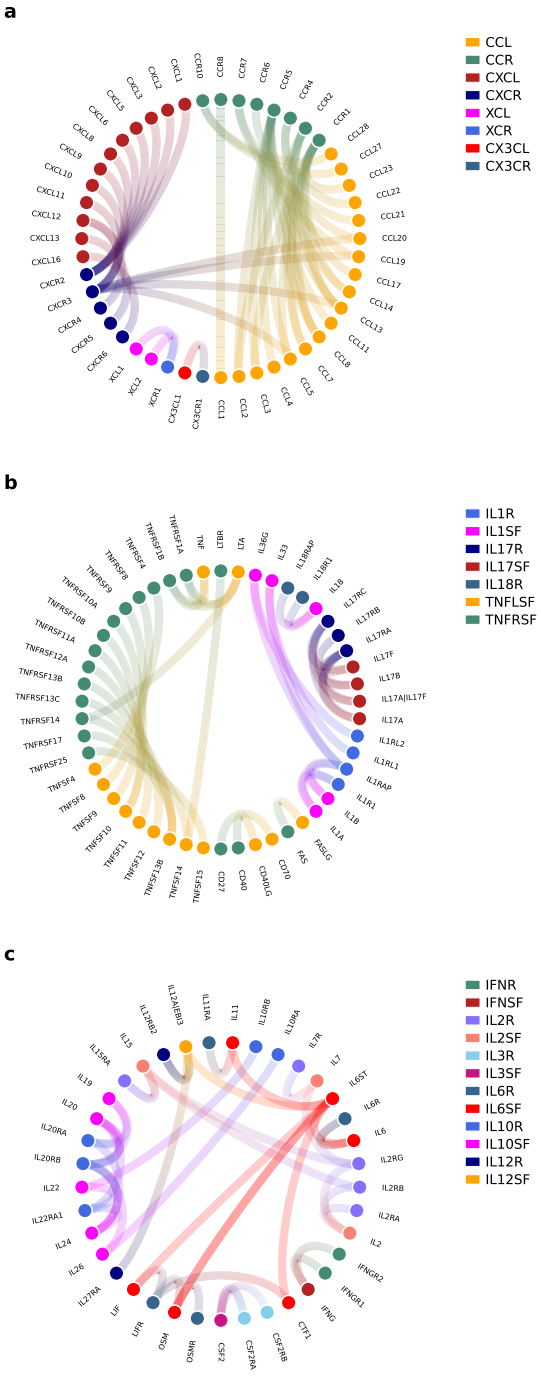

Figure A.6: Psoriasis in skin tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

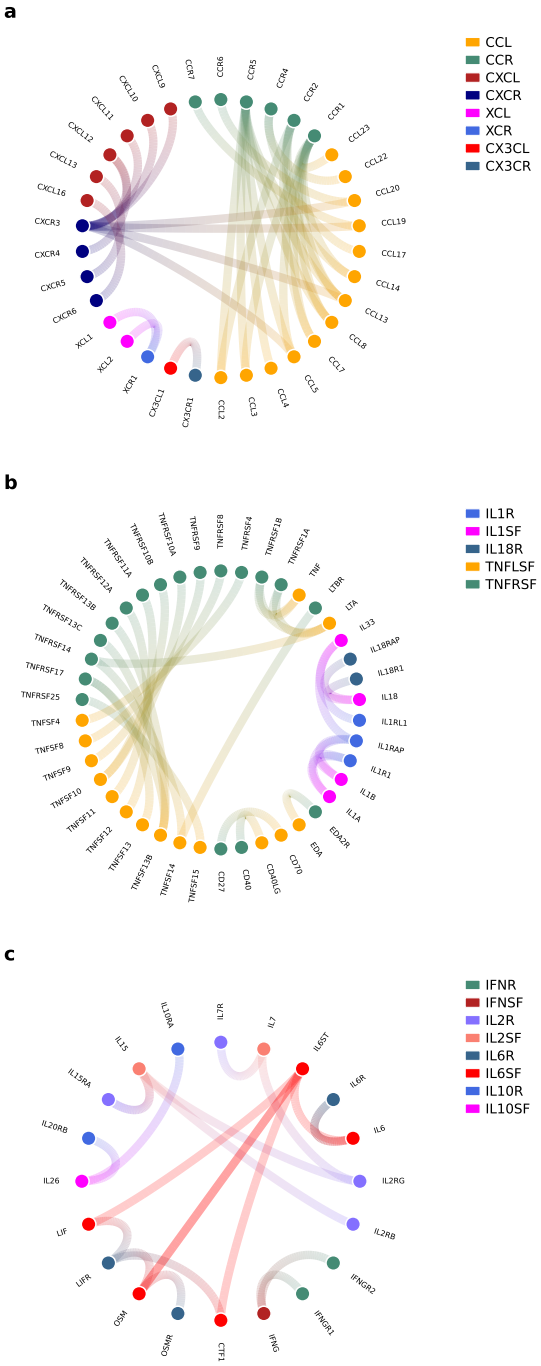

Figure A.7: Psoriatic arthritis in synovial tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

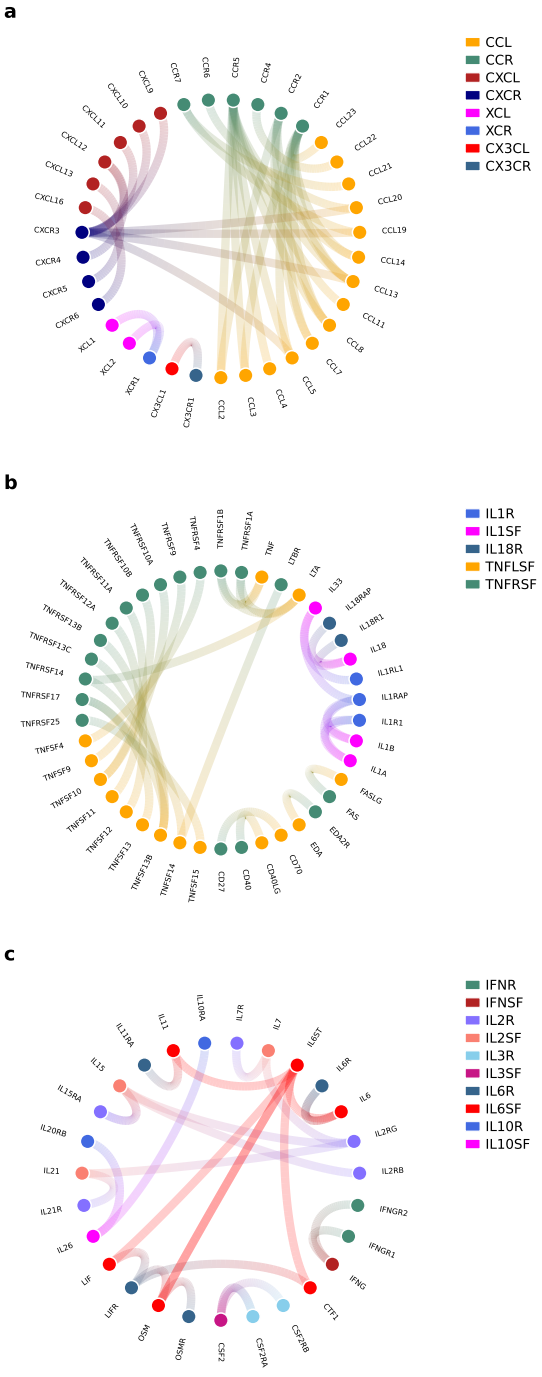

Figure A.8: Rheumatoid arthritis in synovial tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

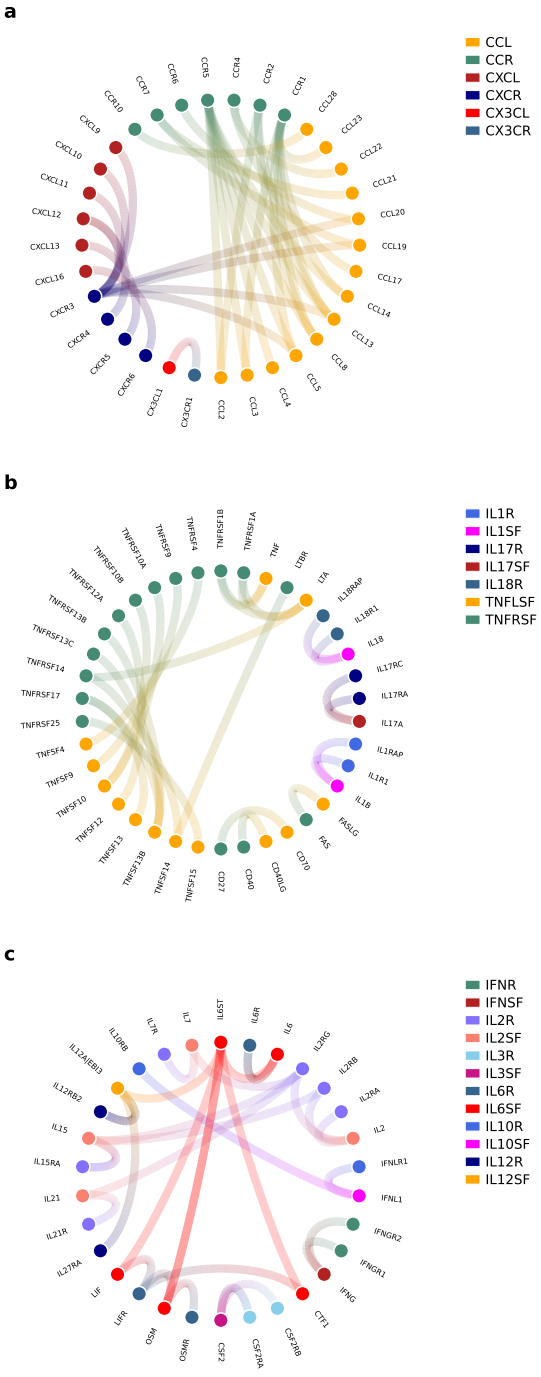

Figure A.9: Sjögren's syndrome in salivary gland tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

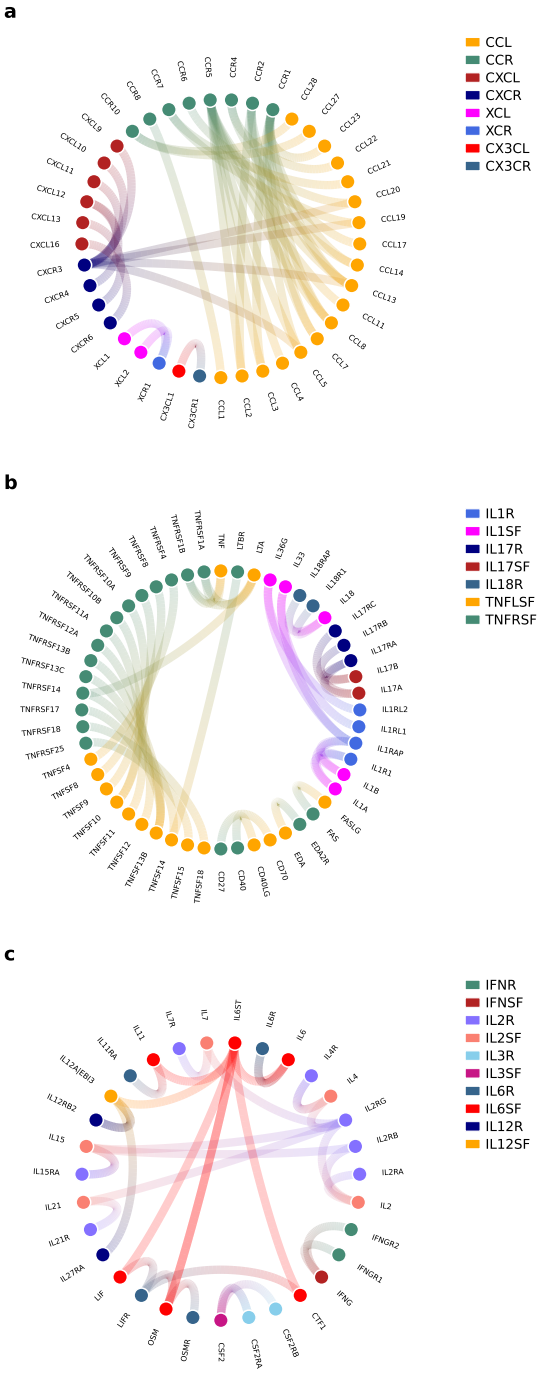

Figure A.10: Systemic lupus erythematosus in skin tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

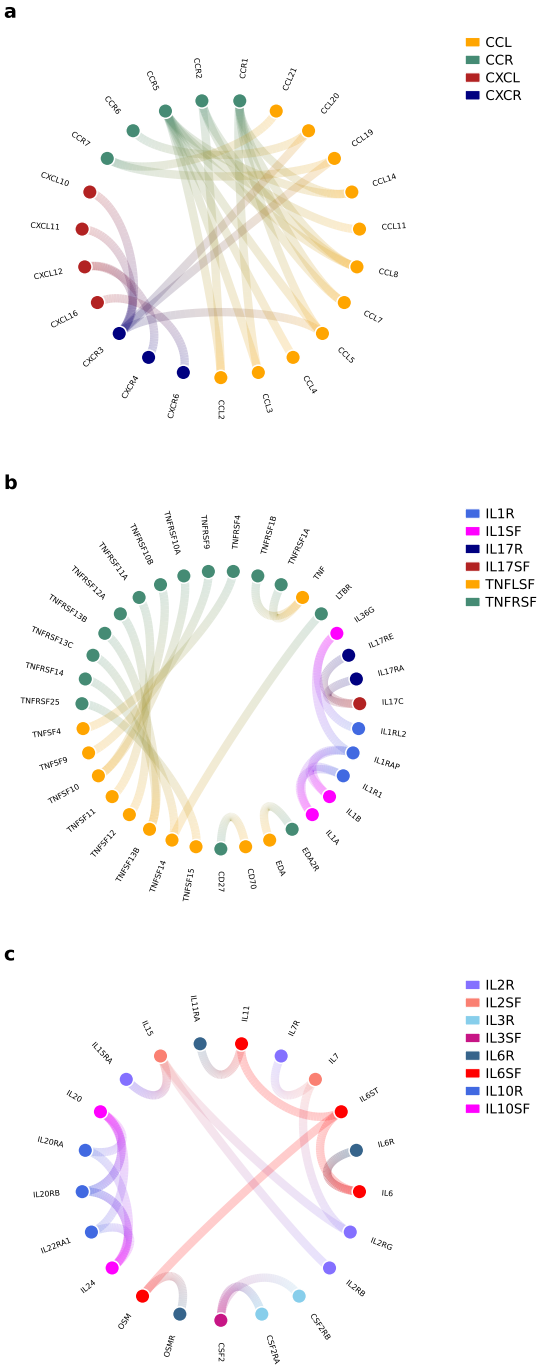

Figure A.11: Type 1 diabetes in pancreatic islet tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.

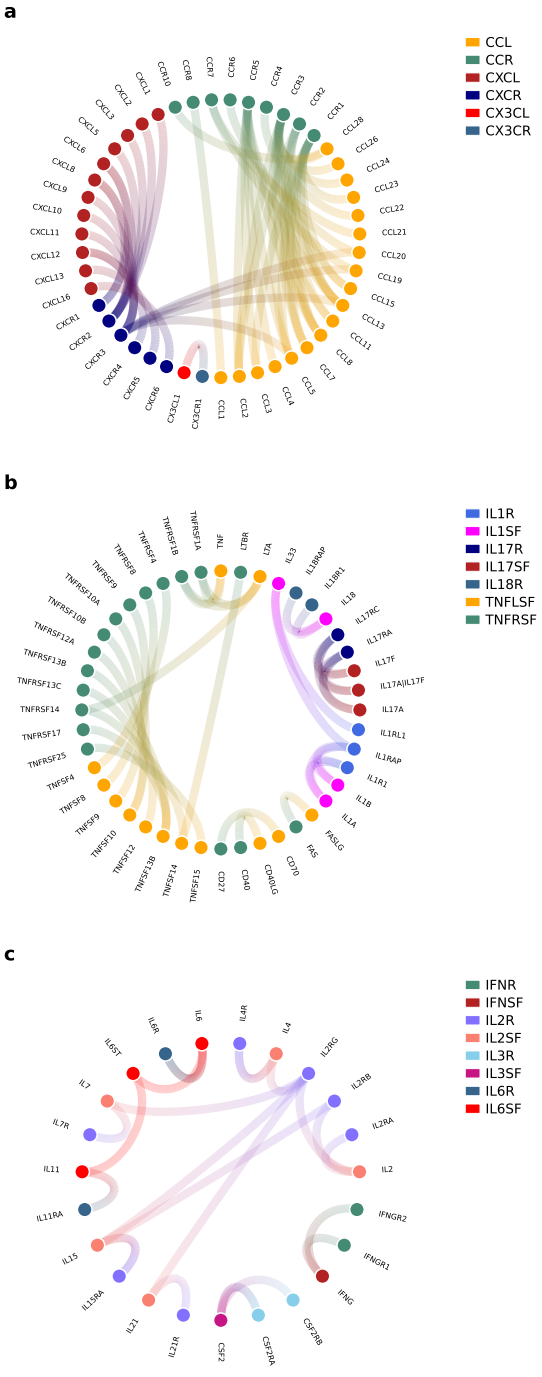

Figure A.12: Ulcerative colitis in colon tissue. **a**, CK:CKR; **b**, NF-κB; **c**, STAT. Nodes represent ligands and receptors; chords represent activating interactions.