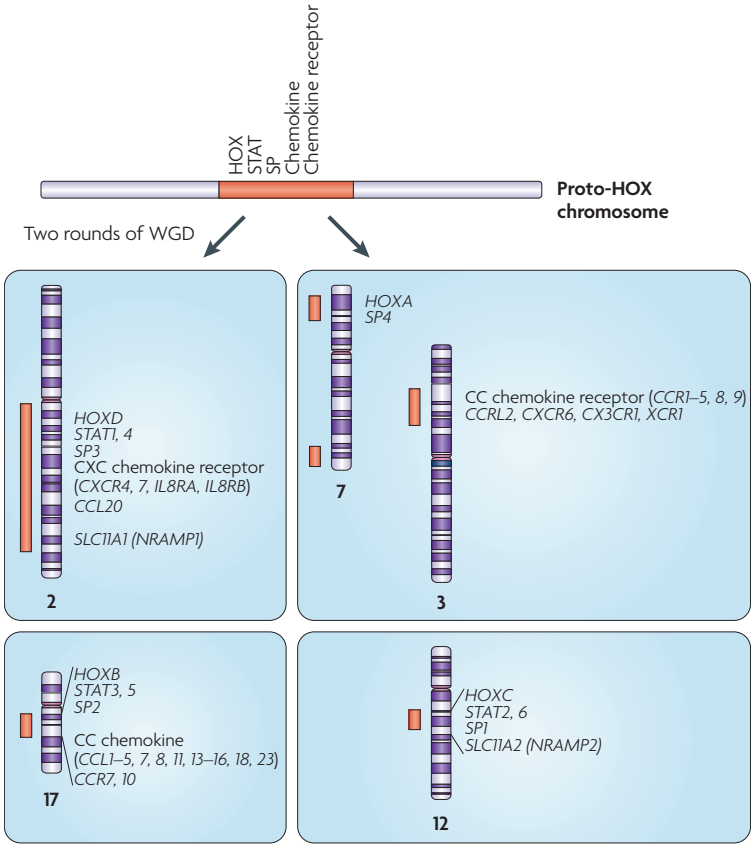

Figure S1 | **The HOX paralogy group.** HOX paralogons are located on human chromosomes 2, 7, 12 and 17. The paralogons on chromosomes 3 and 7 are thought to have been separated by translocations after the two rounds of whole genome duplication (WGD). Jawed vertebrates such as human and shark have four sets of *HOX* clusters whereas the genome of amphioxus contains only one *HOX* cluster[1,2]. Linked to the *HOX* paralogons is a family of genes coding for transcription factors named signal transducers and activators of transcription (STAT), which mediate signal transduction in response to various cytokines. For example, STAT4 and STAT6 mediate transcriptional activation of target genes in response to IL-12 and IL-4, respectively. Their critical roles in adaptive immunity is indicated by the observation that STAT4 deficiency causes a defect in T helper 1 (T_H1) cell development, whereas STAT6 deficiency impairs the development of T helper 2 (T_H2) cells and IL4-dependent Ig class switching. When cytokines are bound to cytokine receptors, the receptor-associated Janus kinases (JAKs) are activated and the activated JAKs phosphorylate STAT proteins, which then form dimers, move to the nucleus and activate transcription of cytokine-responsive genes[3]. Interestingly, all known members of the JAK family (JAK ohnologs) are encoded in the MHC paralogy group, and so are the three of the members of the protein family, known as the protein inhibitors of activated STAT (PIAS), which bind to phosphorylated STAT proteins and prevent their interaction with DNA (Supplementary information S2 (table)). The fact that the key elements of the JAK/STAT pathway are ohnologs located in the prototypical paralogons highlights the effectiveness of WGDs in creating a network of interacting molecules. Other immunologically important genes tightly linked to the *HOX* clusters are those coding for chemokine receptors and chemokines. Not only do chemokines recruit leukocytes including T and B cells to sites of infection, but they regulate physiological migration of lymphocytes to and within various lymphoid tissues. Detailed analysis of chemokines/chemokine receptor genes indicated that they increased their copy number not only by tandem duplication but also by cluster duplication mediated by WGD[4]. Abbreviations: CCL, chemokine CC motif ligand; CCR, chemokine CC motif receptor; CCRL2, chemokine CC motif receptor-like protein 2; CX3CR1, chemokine CX3C motif receptor 1; CXCR, chemokine CXC motif receptor; HOXA~D, homeobox clusters A~D; SLC11A1 and A2, solute carrier family 11, member 1 and member 2; SP, transcription factor SP1; STAT, signal transducer and activator of transcription; and XCR1, chemokine C motif receptor 1

References

1. Garcia-Fernandez, J. The genesis and evolution of homeobox gene clusters. *Nature Rev. Genet.* **6**, 881–892 (2005).
2. Holland, P. W., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. Gene duplications and the origins of vertebrate development. *Development* (Suppl.), 125–133 (1994).
3. Shuai, K., Liu, B. Regulation of JAK–STAT signalling in the immune system. *Nature Rev. Immunol.* **3**, 900–911 (2003).
4. Devries, M. E. *et al.* Defining the origins and evolution of the chemokine/chemokine receptor system. *J. Immunol.* **176**, 401–415 (2006).