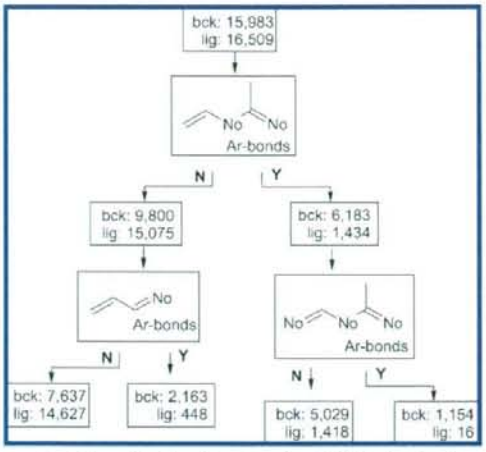


Figure 5. Hierarchical splits for the DIVERSet background compounds compared to GPCR ligands (hGPCR-lig). See legend to Figure 3 for further explanation. Here, it is particularly remarkable that the substructures all contain a double bonded heteroatom. The substructures at the top and right path match the carboxyl and ester groups, which are abundant in the DIVERSet.

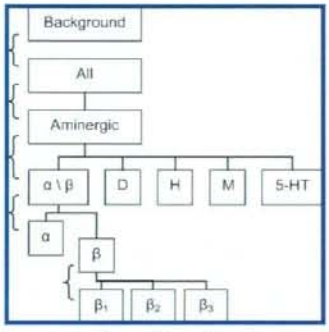


Figure 6. Schematic drawing of the subfamily hierarchy indicating the levels at which substructure analysis was performed (denoted by braces on the left side of the hierarchy). Boxes represent the sets and use the following labels: 'Background' - the ChemBridge DIVERSet, 'All' - total set of GPCR ligands found in GLIDA, 'Aminergic' - all aminergic receptor ligands, '$\alpha \setminus \beta$' - adrenoceptors, 'D' - dopamine receptors, 'H' - histamine, 'M' - muscarinic acetylcholine receptors, '5-HT' - serotonin receptors, 'α' - α-adrenoceptors, 'β' - β-adrenoceptors, and 'β_{1-3}' - β-adrenoceptor subtypes 1 to 3.

is the common motif that accounts for naming of this group as biogenic amine receptors. The high occurrence of these substructures reflects efforts to mimic endogenous ligands by making analogs of these ligands (e.g., isoproterenol based on epinephrine). The opposite analysis was also conducted, yielding the structural features common to GPCR ligands excluding the aminergic ligands. The first, most significant, structural feature was a carbon atom connected to both a single-bonded heteroatom and to a double-bonded heteroatom. In the following positions, this heteroatom was specified as being a nitrogen atom, the second one as an oxygen atom. This reflects the carboxamide motif, found in peptide ligands (MW < 1000), which are part of other, nonaminergic, classes in GLIDA. The second important substructure consisted of two aromatic systems connected by a methylene group or by a single bond.

We continued by analyzing the five major aminergic targets individually against the other four. These five are the adrenoceptors (both alpha- and beta-), the dopamine receptors, the histamine receptors, the muscarinic acetylcholine receptors, and the serotonin receptors. Octopamine and trace amine receptors were not included due to scarce ligand information. For each analysis, the size of the aminergic control group was different due to the removal of duplicate entries, i.e. compounds that bind to more than one class. Although substructures found for the control group may be common to multiple GPCR targets, these are different from privileged substructures. Privileged substructures are discrete fragments, often scaffolds, found in one or more ligands for more than one target in the family.[47] Our analysis considers all possible substructures and yields only the most frequent substructures among the targets.

Adrenoceptor Ligands. An important feature of the adrenergic receptor ligands vs all other aminergic ligands is a substructure consisting of two heteroatoms connected by an ethyl group (Figure 8). The first heteroatom of this substructure is an oxygen atom specified as a hydroxy group, and the second is a nitrogen atom with a single hydrogen atom attached, meaning that this nitrogen is secondary. This chemical signature is representative for the motifs found in both β-adrenoceptor agonists and antagonists. An example containing this substructure is metoprolol, a β_1 antagonist (beta-blocker) used to treat hypertension. The second example substructure for motif I in Figure 8 has no atom specifiers for the heteroatoms, which means that this substructure also overlaps with the 1,2 diaminoethane substructure. A search for adrenoceptor (ant)agonists that have this substructure and not the hydroxyethylamine returned 58 hits, most of them specified as α-adrenoceptor ligands in the database (second example structure of motif I). Note that both aforementioned substructures in the query had heteroatoms with one explicit hydrogen atom. At lower positions, the hydroxyethylamine motif reappears bonded to an aromatic system at the carbon atom that has the hydroxyl group attached. This is an exclusive element in β-adrenoceptor agonists. The substructures are essentially all part of the example substructure given for motif II which is found in 27% of the aminergic ligands. An example drug that has this motif is terbutaline, a β_2-adrenoceptor agonist used in the treatment of asthma. Substructures found less frequently in adrenergic ligands compared to aminergic ligands consisted of a nitrogen atom substituted at two or three positions, some as part of a largely saturated five- or six-membered ring, as found in e.g., apomorphine, a dopamine receptor ligand.

Alpha- and Beta-Adrenoceptor Ligands. We further examined the adrenergic receptor ligands, where we distinguished between α- and β-adrenoceptors. The most significant features specific for the α-adrenoceptor ligands (Figure 9) consist of a nitrogen atom substituted at three positions with methyl and ethyl groups (73% of ligands). One ethyl group can be connected to an aromatic system (33%) or to a heteroatom that is connected to an aromatic system (29%). An example drug containing this substructure is phenoxybenzamine, an α_1-adrenoceptor antagonist used in the treatment of hypertension. The most significant substructures specific for β-adrenoceptor ligands (Figure 10) were all based on the 1-(ethylamino)propan-2-ol moiety (86% of ligands).

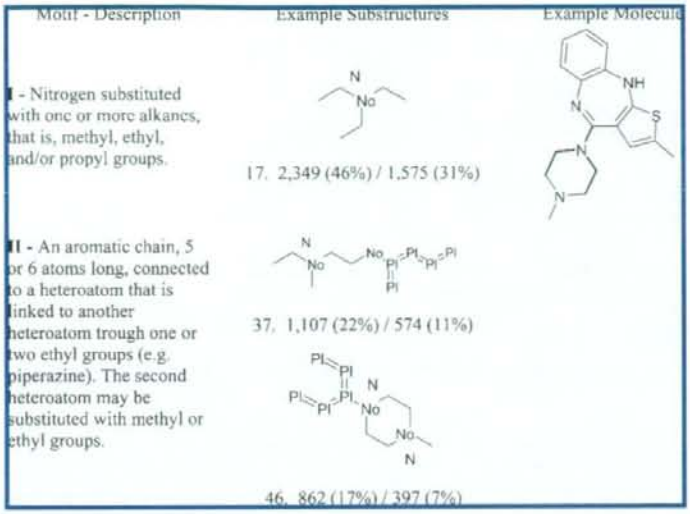


Figure 7. Common motif and example substructures for most significant substructures of aminergic ligands compared against all other GPCR ligands (GLIDA, 5k sampled), in planar ring systems representation. The 'Motif' number (Roman number, in bold) indicates the number of the found motif or structural theme. The motif number is followed by a short description, and one or more example substructures are provided. Below each example substructure, the position, occurrence in the active set (absolute and percentage), and occurrence in the control set (absolute and percentage) are listed. See the Materials and Methods section for further explanation about the representation of the substructures. For some motifs, an example molecule from the same class is provided, with the example substructure overlaid in bold. Here, an example drug containing motif I is olanzapine, which is used to treat schizophrenia, acting on dopamine D_1, D_2, and serotonin 5-HT_2 receptors (taken from ref 51).

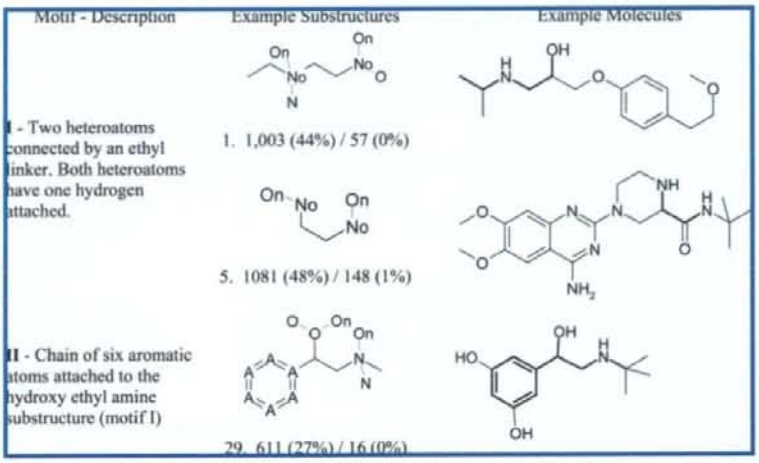


Figure 8. Common motif and example substructures for most significant substructures of the adrenoceptors ligands, in aromatic atoms and bonds representation. See the legend of Figure 7 for further explanation. The first example structure (for motif I) is metoprolol, a β_1-adrenoceptor antagonist (beta-blocker) used to treat hypertension (taken from ref 51). The second example structure for motif I is 4-(4-amino-6,7-dimethoxyquinazolin-2-yl)-N-tert-butylpiperazine-2-carboxamide, an α-adrenoceptor ligand and prazosin derivative found in GLIDA.[33] The example structure for motif II is terbutaline, a β_2-adrenoceptor agonist used in the treatment of asthma.

An example drug containing this substructure is propranolol, a nonselective beta-blocker, used in the treatment of hypertension. The most significant substructures specific for the β_1-adrenoceptor were all parts of a methylaminopropane substructure (81% of ligands). Here it should be noted that commercially available β_2-adrenoceptor ligands are *agonist*-shaving a structure such as terbutaline (Figure 8, first example motif II), whereas β_1-adrenoceptor ligands are mostly antagonists such as metoprolol (Figure 8, first example motif I). The most significant *avoiding* substructure for β_1-adrenoceptor ligands (50% of ligands), which at the same time occurs in β_2- and β_3-adrenoceptor ligands, consisted of an aromatic chain linked by an ethyl group to nitrogen that was linked by an ethyl group to an oxygen.

Dopamine Receptor Ligands. For the dopamine receptor ligands, two types of specific substructures were identified (Figure 11). The first substructure (in 30% of the ligands) consists of a chain of 4 to 5 aromatic atoms, connected to a nitrogen atom through a single carbon atom. This nitrogen is tertiary, as it is substituted with either two ethyl groups or one methyl and one ethyl group. The second substructure (12% of the ligands) consists of two aromatic chains of five or six atoms long that are linked through a heteroatom connected to N-methylethyleneamine, e.g. an N-methyleth-

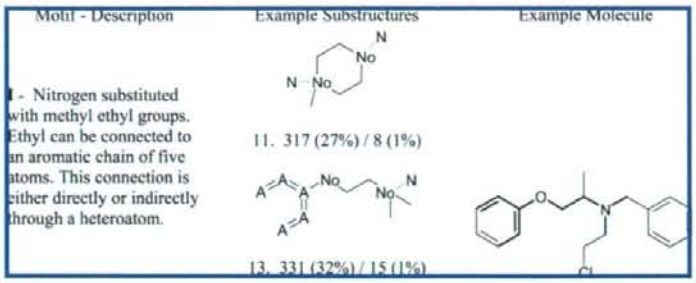


Figure 9. Common motif and example substructures for most significant substructures of the α-adrenoceptors ligands versus β-adrenoceptor ligands, in aromatic atoms and bonds representation. An example is phenoxybenzamine, a α_1-receptor antagonist used to treat hypertension. See the legend of Figure 7 for further explanation.

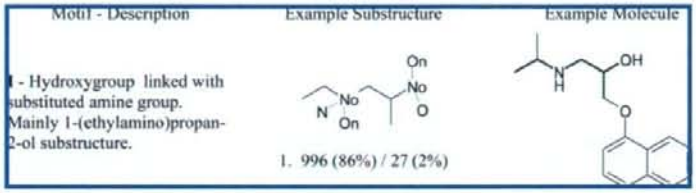


Figure 10. Common motif and example substructures for most significant substructures of the β-adrenoceptor ligands versus α-adrenoceptors ligands, in aromatic bonds representation. An example drug containing this substructure is propranolol, a nonselective β-adrenoceptor antagonist (beta-blocker). See the legend of Figure 7 for further explanation.

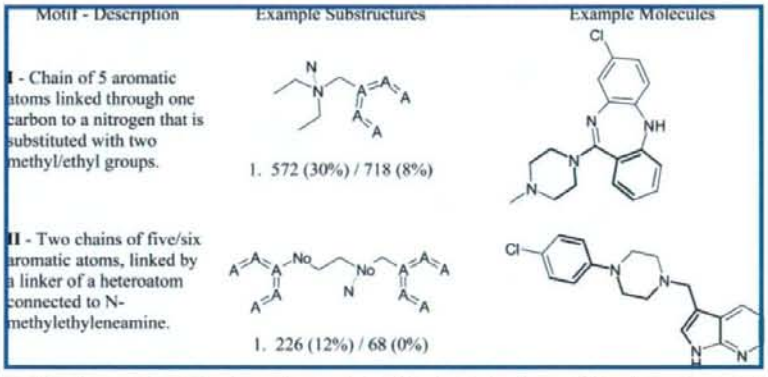


Figure 11. Common motif and example substructures for most significant substructures of the dopamine receptor ligands, in aromatic atoms and bonds representation. An example drug that has motif I is clozapine, an antipsychotic agent used in the treatment of schizophrenia.[52] Another example for motif I and also for motif II is compound L-745,870, a selective dopamine D_4 receptor antagonist.[53] See the legend of Figure 7 for further explanation.

ylenediamine linker. The terminal nitrogen of this linker may be substituted with an ethyl group. In both example molecules in Figure 11, the substructures overlap with the piperazine ring. The fact that the most significant substructures do not 'use' the entire piperazine moiety suggests that variations on the piperazine theme are possible when designing dopaminergic drugs. Similarly, the aromatic chains in motifs I and II overlap with several types of aromatic systems, e.g. five-membered or six-membered rings, containing either carbon or heteroatoms. This implies that aromaticity is the important feature and not so much the type of ring system that is used. Again, this finding offers further options for drug design.

Histamine Receptor Ligands. The most common motif (almost 50%) specific for histamine receptors is a chain of five aromatic atoms (Figure 12), where one or two aromatic atoms are specified as nitrogen atoms. These nitrogen atoms are separated by one aromatic atom; in some cases, one of

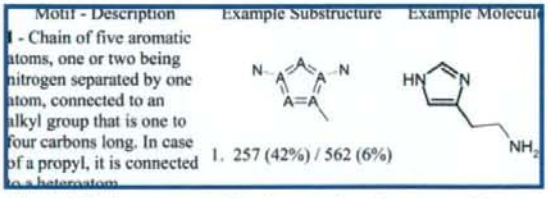


Figure 12. Common motif and example substructures for most significant substructures of the histamine receptor ligands, in aromatic atoms and bonds representation. Histamine is provided as an example molecule containing this motif. See the legend of Figure 7 for further explanation.

the other neighboring aromatic atoms has an ethyl group attached. The majority of the significant substructures are chains, and actual ring closures, forming e.g. the imidazole ring as in histamine, are scarce. This seems counterintuitive at first sight, since the five-membered aromatic heterocycles are among the most obvious features when visually inspect-

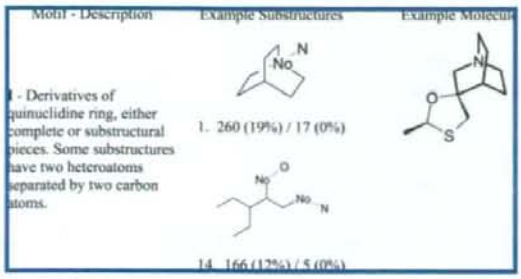


Figure 13. Common motif and example substructures for most significant substructures of the muscarinic acetylcholine receptor ligands, in aromatic bonds representation. An example molecule containing the quinuclidine ring is cevimeline, a muscarinic M_3 receptor agonist. See the legend of Figure 7 for further explanation.

ing the set. Although a common theme, the heterocycles in histamine receptor ligands all differ in size, ring-fusions, and heteroatoms. By considering substructures instead of complete ring fragments, it was thus possible to find structural similarities that have a much higher support among the ligands. This causes the high occurrence of the 'aromatic chains', since it is the most common feature among the diverse heterocycles.

Muscarinic Acetylcholine Receptor Ligands. Substructures of derivatives of the quinuclidine ring are the most common substructures for the muscarinic acetylcholine receptor ligands (Figure 13). It occurs in 19% of ligands for this class compared to 0% in other aminergic ligands. A second heteroatom may be attached, separated by two carbon atoms from the nitrogen. A typical example is civemeline, a muscarinic M_3 receptor agonist (Figure 13, first example). In this case, the second heteroatom would be the oxygen atom.

Serotonin Receptor Ligands. For the serotonin receptor ligands (Figure 14), the most specific substructure resembles the shape of the 2-ethylindole moiety (31% of ligands) that forms the core of serotonin, although a label specifying the nitrogen atom is missing. This is because this substructure covers the largest set of serotonin ligands (without becoming too general). In some cases, the ethyl group is attached to C3 rather than N1 of the core, which means that either the ethyl group or the atom specifier is not part of the substructure. The nitrogen atom can also be replaced by other heteroatoms or be absent in scaffolds that consist only of carbons, forming a planar ring system. This ring system consists of one aromatic ring instead of two. In fact, examples of all three cases were found among the ligands. At lower positions in the lists (position 22 in Table 16, Supporting Information) the same substructure (25% of ligands) is found with the nitrogen atom specifier. Examples containing this substructure are the endogenous ligand serotonin and the triptan antimigraine drugs such as sumatriptan (Figure 14).

Aminergic Receptor Ligands. When comparing the aminergic subgroups against all other aminergic ligands, the most significant features of the aminergic supergroup always have a low occurrence. This low occurrence is probably because these features are actually the features of one of the other subgroups of this class. For instance, the *avoiding* substructures for the serotonin receptor ligands are, among others, motif I from the adrenoceptor ligands (Figure 8). Therefore, these features can be considered substructures to avoid since they indicate possible side effects on other aminergic receptors. This is a more generalized way of a so-called antitarget analysis, to avoid GPCR-mediated side effects.[48] For dopamine, histamine, and serotonin receptor ligands, the most significant substructures of the other aminergic ligands (the *avoiding* substructures) are dominated by a motif of two heteroatoms connected by an ethyl linker. Since this resembles the substructures found for the (β) adrenoceptors (Figure 8), in both frequency and shape, this class probably dominates the avoiding substructure lists of the other four classes.

General Observations. In the following, we will discuss the representations and substructure selection criteria employed and their likely influence on the results obtained. First, the extraction of substructures discards any geometric information such as bond orientation. This loss of information may be appreciated, however, as it is beneficial for extracting more 'abstract' features in molecules. For instance, opposite *cis*−*trans* isomers may contribute to the same double bond in a substructure. Similarly, a chain of aromatic bonds may be part of one or multiple fused ring systems. Chirality is also lost in our approach, an issue that holds for all substructure search methods. Inclusion of 3D-conformational aspects in substructure searching is an open area for further research. Second, the p-value was used to sort the substructures according to significance. However, this value is very small for the top findings, and the differences in significance of the substructures are small. Therefore, choosing the most significant substructure to split the set is arbitrary; all substructures at the top of the occurrence lists would be a very good choice. Not only the significance of the finding is important but also what the finding predicts. It might therefore be better to focus less on the p-value and more on the ratio of retrieval of GPCR ligands and background compounds. In the end, a scientist might be more interested in the percentage of GPCR ligands that can be identified

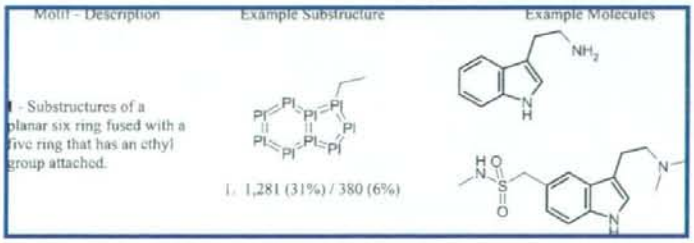


Figure 14. Common motif and example substructures for most significant substructures of the serotonin ligands, in planar ring systems representation. An example substructure is the endogenous ligand serotonin (above), or the antimigraine drug sumatriptan (below). See the legend of Figure 7 for further explanation.

from a set of molecules. Moreover, our approach focuses on selective substructures only; substructures that occur frequently in both sets are discarded. Although not selective, these substructures are still common structural features for GPCR ligands. Therefore, for the design of GPCR screening libraries, we also analyzed the features common in GPCR ligands in general, which were selected based on a minimum number of atoms/bonds (to remove trivial fragments) and frequency. A list of frequent, nonselective substructures is provided in Table 22, Supporting Information. This table lists all frequent (≥60% of molecules) substructures of more than six atoms that are found in GLIDA (5k sampled, normal representation). Some of the substructures (such as no. 23) can be associated easily with ligands of the aminergic receptor class, while other substructures (such as no. 1) seem to describe a more generic pattern of GPCR ligands. While this table will not be analyzed further by us at the current stage, fragments of the above type should be beneficial to be included in the design of GPCR libraries of any of the subtypes discussed in this work. As a final point, almost all significant findings were found when we used elaborate chemical representations, mainly with 'aromatic atom and bond types'. This suggests that this representation might be best for the application domain (i.e., GPCR activity prediction). Apparently, elaborate chemical representations add substantial value when searching for structural features typical for active compounds. Suggestions for further research would therefore be to extend the types of representations used, for instance by encoding the electronic properties of a molecule (for example, see ref 49).

CONCLUSION

In this study, we analyzed frequently occurring substructures in GPCR ligands by comparing these with various control groups of compounds. Our analysis is complementary to employing privileged structures in ligand design,[50] in that it is not restricted to existing scaffold structures. It therefore offers further opportunities for introducing novelty in new chemical entities. We used different chemical representations for the molecules under consideration. As a result, we derived generalized substructural features for both ligands and control groups. The substructures found in the background set reflect the use of simple reactions that may have been employed to construct the library, for instance, the ester and carboxamide groups. In the GPCR ligand group, we found common as well as 'novel' substructures. First of all, our analysis identified well-known motifs (e.g., the side chain in β-adrenoceptor antagonists), which we considered a validation of our approach. In fact, the butylamine substructure (often linked to an aromatic moiety) occurred in 74% of the GPCR ligands compared to 26% of the background control group. Second, new structural patterns were also found, which may help medicinal chemists in their design efforts. As a typical example, we found fused 5:6 bicyclic ring systems in serotonergic ligands. These were identified in the so-called planar representation, indicating that aromaticity is not essential for both rings and that the precise location and nature of a heteroatom in the bicyclic core is not fixed. Indeed, the use of elaborate chemical representation gave the best, i.e. the most significant, description of the structural features that are important for a (sub)class of GPCR ligands.

ACKNOWLEDGMENT

The authors thank Jeroen Kazius and Siegfried Nijssen for helpful discussions. This study was (partially) performed within the framework of the Dutch Top Institute Pharma, project number: D1-105.

Supporting Information Available: Lists of most significant substructures per compound set. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–2893.
(2) Bemis, G. W.; Murcko, M. A. Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42* (25), 5095–5099.
(3) Xue, L.; Bajorath, J. Distribution of Molecular Scaffolds and R-Groups Isolated from Large Compound Databases. *J. Mol. Model.* **1999**, *5* (5), 97–102.
(4) Xu, J. A New Approach to Finding Natural Chemical Structure Classes. *J. Med. Chem.* **2002**, *45* (24), 5311–5320.
(5) Nilakantan, R.; Nunn, D. S.; Greenblatt, L.; Walker, G.; Haraki, K.; Mobilio, D. A Family of Ring System-Based Structural Fragments for Use in Structure-Activity Studies: Database Mining and Recursive Partitioning. *J. Chem. Inf. Model.* **2006**, *46* (3), 1069–1077.
(6) Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Van Aken, K.; Janssen, P. A. J. SYNOPSIS: SYNthesize and OPtimize System in Silico. *J. Med. Chem.* **2003**, *46* (13), 2765–2773.
(7) Todd, M. H. Computer-aided organic synthesis. *Chem. Soc. Rev.* **2005**, *34* (3), 247–66.
(8) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1708–1718.
(9) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (1), 170–178.
(10) Lameijer, E. W.; Kok, J. N.; Back, T.; IJzerman, A. P. Mining a Chemical Database for Fragment Co-occurrence: Discovery of "Chemical Clichés". *J. Chem. Inf. Model.* **2006**, *46* (2), 553–562.
(11) Lipkus, A. H.; Yuan, Q.; Lucas, K. A.; Funk, S. A.; Bartelt, W. F.; Schenck, R. J.; Trippe, A. J. Structural Diversity of Organic Chemistry. A Scaffold Analysis of the CAS Registry. *J. Org. Chem.* **2008**, *73* (12), 4443–4451.
(12) Aronov, A. M.; Bemis, G. W. A minimalist approach to fragment-based ligand design using common rings and linkers: application to kinase inhibitors. *Proteins* **2004**, *57* (1), 36–50.
(13) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2* (22), 3204–3218.
(14) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273–1280.
(15) Glen, R. C.; Bender, A.; Arnby, C. H.; Carlsson, L.; Boyer, S.; Smith, J. Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME. *IDrugs* **2006**, *9* (3), 199–204.
(16) Batista, J.; Godden, J. W.; Bajorath, J. r. Assessment of Molecular Similarity from the Analysis of Randomly Generated Structural Fragment Populations. *J. Chem. Inf. Model.* **2006**, *46* (5), 1937–1944.
(17) Batista, J.; Bajorath, J. Chemical Database Mining through Entropy-Based Molecular Similarity Assessment of Randomly Generated Structural Fragment Populations. *J. Chem. Inf. Model.* **2007**, *47* (1), 59–68.
(18) Sheridan, R. P.; Miller, M. D. A Method for Visualizing Recurrent Topological Substructures in Sets of Active Molecules. *J. Chem. Inf. Model.* **1998**, *38* (5), 915–924.
(19) Raymond, J. W.; Willett, P. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J. Comput.-Aided Mol. Des.* **2002**, *16* (7), 521–533.
(20) *Scitegic Pipeline Pilot, 6.1.5.0 Student Edition*; Accelrys, Inc.: San Diego, CA, 2007.
(21) *ClassPharmer, 4.5*; Simulations Plus, Inc.: Lancaster, CA, 2008.
(22) Wörlein, M.; Meinl, T.; Fischer, I.; Philippsen, M. A Quantitative Comparison of the Subgraph Miners MoFa, gSpan, FFSM, and Gaston*Knowledge Discovery in Databases: PKDD 2005*; 2005; pp 392−403.

(23) Engkvist, O.; Wrede, P.; Rester, U. Prediction of CNS Activity of Compound Libraries Using Substructure Analysis. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (1), 155–160.
(24) Borgelt, C.; Berthold, M. R. Mining Molecular Fragments: Finding Relevant Substructures of Molecules. In *Data Mining*, Proceedings of the 2002 IEEE International Conference on Data Mining, IEEE Computer Society: pp 51−58.
(25) Barratt, M. D.; Rodford, R. A. The computational prediction of toxicity. *Curr. Opin. Chem. Biol.* **2001**, *5* (4), 383–388.
(26) Nijssen, S.; Kok, J. N. A quickstart in frequent structure mining can make a difference. In *Conference on Knowledge Discovery in Data*, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. Ronny, K., Gehrke, J., DuMouchel, W., Ghosh, J., Eds.; ACM Press: New York, U.S.A.: pp 647−652.
(27) Nijssen, S. *MULTI GASTON GrAph, Sequences and Tree ExtractiON algorithm, version 0.2*; Leiden Institute of Advanced Computer Science, Leiden University: Leiden, The Netherlands, 2004.
(28) Kazius, J.; Nijssen, S.; Kok, J.; Bäck, T.; IJzerman, A. P. Substructure Mining Using Elaborate Chemical Representation. *J. Chem. Inf. Model.* **2006**, *46* (2), 597–605.
(29) Heiko, H.; Christian, B.; Michael, R. B. *Large Scale Mining of Molecular Fragments with Wildcards*. Procedings of the 5th International Symposium on Intelligent Data Analysis, IOS Press: pp 495−504.
(30) Meinl, T.; Borgelt, C.; Berthold, M. R. *Mining Fragments with Fuzzy Chains in Molecular Databases*, Proceeding of the 2nd International Workshop on Mining Graphs, Trees and Sequences, Pisa, Italy, pp 49−60.
(31) Balakin, K. V.; Tkachenko, S. E.; Lang, S. A.; Okun, I.; Ivashchenko, A. A.; Savchuk, N. P. Property-Based Design of GPCR-Targeted Library. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1332–1342.
(32) Schnur, D. M.; Hermsmeier, M. A.; Tebben, A. J. Are Target-Family-Privileged Substructures Truly Privileged. *J. Med. Chem.* **2006**, *49* (6), 2000–2009.
(33) Okuno, Y.; Tamon, A.; Yabuuchi, H.; Niijima, S.; Minowa, Y.; Tonomura, K.; Kunimoto, R.; Feng, C. GLIDA: GPCR ligand database for chemical genomics drug discovery database and tools update. *Nucleic Acids Res.* **2008**, *36* (suppl_1), D907−912.
(34) hGPCR - lig. http://bioinfo-pharma.u-strasbg.fr:8080/hGPCRLig/index.jsp (accessed March 20, 2007).
(35) Wheeler, D. L.; Barrett, T.; Benson, D. A.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Edgar, R.; Federhen, S.; Feolo, M.; Geer, L. Y.; Helmberg, W.; Kapustin, Y.; Khovayko, O.; Landsman, D.; Lipman, D. J.; Madden, T. L.; Maglott, D. R.; Miller, V.; Ostell, J.; Pruitt, K. D.; Schuler, G. D.; Shumway, M.; Sequeira, E.; Sherry, S. T.; Sirotkin, K.; Souvorov, A.; Starchenko, G.; Tatusov, R. L.; Tatusova, T. A.; Wagner, L.; Yaschenko, E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2008**, *36*, D13−D21.
(36) Roth, B. L.; Lopez, E.; Beischel, S.; Westkaemper, R. B.; Evans, J. M. Screening the receptorome to discover the molecular targets for plant-derived psychoactive compounds: a novel approach for CNS drug discovery. *Pharmacol. Ther.* **2004**, *102* (2), 99–110.
(37) *DIVERSet*; ChemBridge Corp.: San Diego, CA, 2006.
(38) IUPHAR RECEPTOR DATABASE. www.iuphar-db.org (accessed May 9, 2007).
(39) Foord, S. M.; Bonner, T. I.; Neubig, R. R.; Rosser, E. M.; Pin, J.-P.; Davenport, A. P.; Spedding, M.; Harmar, A. J. International Union of Pharmacology. XLVI. G Protein-Coupled Receptor List. *Pharmacol. Rev.* **2005**, *57* (2), 279–288.
(40) Horn, F.; Bettler, E.; Oliveira, L.; Campagne, F.; Cohen, F. E.; Vriend, G. GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.* **2003**, *31* (1), 294–297.
(41) *JChem Standardizer, 3.2.11*; ChemAxon Kft.: Budapest, Hungary, 2007.
(42) Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. *J. Med. Chem.* **2005**, *48* (1), 312–320.
(43) *ISIS/Draw, 2.5*; MDL Information Systems, Inc.: San Leandro, CA, 2002.
(44) Strader, C. D.; Sigal, I. S.; Candelore, M. R.; Rands, E.; Hill, W. S.; Dixon, R. A. Conserved aspartic acid residues 79 and 113 of the β-adrenergic receptor have different roles in receptor function. *J. Biol. Chem.* **1988**, *263* (21), 10267–10271.
(45) Lindner, M. D. Clinical attrition due to biased preclinical assessments of potential efficacy. *Pharmacol. Ther.* **2007**, *115* (1), 148–175.
(46) Feher, M.; Schmidt, J. M. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (1), 218–27.
(47) Bondensgaard, K.; Ankersen, M.; Thogersen, H.; Hansen, B. S.; Wulff, B. S.; Bywater, R. P. Recognition of Privileged Structures by G-Protein Coupled Receptors. *J. Med. Chem.* **2004**, *47* (4), 888–899.
(48) Klabunde, T.; Evers, A. GPCR Antitarget Modeling: Pharmacophore Models for Biogenic Amine Binding GPCRs to Avoid GPCR-Mediated Side Effects. *ChemBioChem* **2005**, *6* (5), 876–889.
(49) Marin, R. M.; Aguirre, N. F.; Daza, E. E. Graph Theoretical Similarity Approach To Compare Molecular Electrostatic Potentials. *J. Chem. Inf. Model.* **2008**, *48* (1), 109–118.
(50) Jacoby, E.; Fauchère, J.-L.; Raimbaud, E.; Ollivier, S.; Michel, A.; Spedding, M. A Three Binding Site Hypothesis for the Interaction of Ligands with Monoamine G Protein-coupled Receptors: Implications for Combinatorial Ligand Design. *Quant. Struct.-Act. Relat.* **1999**, *18* (6), 561–572.
(51) Klabunde, T.; Hessler, G. Drug Design Strategies for Targeting G-Protein-Coupled Receptors. *ChemBioChem* **2002**, *3* (10), 928–944.
(52) Meltzer, H. Y. Treatment-resistant schizophrenia--the role of clozapine. *Curr. Med. Res. Opin.* **1997**, *14* (1), 1–20.
(53) Bristow, L. J.; Kramer, M. S.; Kulagowski, J.; Patel, S.; Ragan, C. I.; Seabrook, G. R. Schizophrenia and L-745, 870, a novel dopamine D4 receptor antagonist. *Trends Pharmacol. Res.* **1997**, *18* (6), 186–188.
(54) van der Horst, E.; IJzerman, A. P. Computational Approaches to Fragment and Substructure Discovery and Evaluation. In *Fragment-Based Drug Discovery: a Practical Approach*, first edition; Zartler, E. R., Shapiro, M. J., Eds.; John Wiley and Sons, Ltd.: Chichester, West-Sussex, United Kingdom, 2008; pp 199–222.

CI8003896

BMC Genomics

BioMed Central

Research article

Open Access

Characterization of gene expression profiles for different types of mast cells pooled from mouse stomach subregions by an RNA amplification method

Soken Tsuchiya[1], Yuki Tachida[1], Eri Segi-Nishida[1,2], Yasushi Okuno[2,3], Shigero Tamba[1], Gozoh Tsujimoto[4], Satoshi Tanaka[1,5] and Yukihiko Sugimoto*[1]

Address: [1]Department of Physiological Chemistry, Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan, [2]Department of Systems Bioscience for Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan, [3]Department of PharmacoInformatics, Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan, [4]Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan and [5]Department of Immunobiology, School of Pharmacy and Pharmaceutical Sciences, Mukogawa Women's University, Nishinomiya, Hyogo 663-8179, Japan

Email: Soken Tsuchiya - Soken.Tsuchiya@f13.mbox.media.kyoto-u.ac.jp; Yuki Tachida - yuki.tachida@090014.mbox.media.kyoto-u.ac.jp; Eri Segi-Nishida - eri.segi.nishida@pharm.kyoto-u.ac.jp; Yasushi Okuno - okuno@pharm.kyoto-u.ac.jp; Shigero Tamba - livegym-tmb@m6.gyao.ne.jp; Gozoh Tsujimoto - gtsuji@pharm.kyoto-u.ac.jp; Satoshi Tanaka - s_tanaka@mukogawa-u.ac.jp; Yukihiko Sugimoto* - ysugimot@pharm.kyoto-u.ac.jp

* Corresponding author

Published: 20 January 2009

BMC Genomics 2009, **10**:35 doi:10.1186/1471-2164-10-35

Received: 9 July 2008
Accepted: 20 January 2009

This article is available from: http://www.biomedcentral.com/1471-2164/10/35

Abstract

Background: Mast cells (MCs) play pivotal roles in allergy and innate immunity and consist of heterogenous subclasses. However, the molecular basis determining the different characteristics of these multiple MC subclasses remains unclear.

Results: To approach this, we developed a method of RNA extraction/amplification for intact *in vivo* MCs pooled from frozen tissue sections, which enabled us to obtain the global gene expression pattern of pooled MCs belonging to the same subclass. MCs were isolated from the submucosa (sMCs) and mucosa (mMCs) of mouse stomach sections, respectively, 15 cells were pooled, and their RNA was extracted, amplified and subjected to microarray analysis. Known marker genes specific for mMCs and sMCs showed expected expression trends, indicating accuracy of the analysis.

We identified 1,272 genes showing significantly different expression levels between sMCs and mMCs, and classified them into clusters on the basis of similarity of their expression profiles compared with bone marrow-derived MCs, which are the cultured MCs with so-called 'immature' properties. Among them, we found that several key genes such as *Notch4* had sMC-biased expression and *Ptgr1* had mMC-biased expression. Furthermore, there is a difference in the expression of several genes including extracellular matrix protein components, adhesion molecules, and cytoskeletal proteins between the two MC subclasses, which may reflect functional adaptation of each MC to the mucosal or submucosal environment in the stomach.

Conclusion: By using the method of RNA amplification from pooled intact MCs, we characterized the distinct gene expression profiles of sMCs and mMCs in the mouse stomach. Our findings offer insight into possible unidentified properties specific for each MC subclass.

Background

Mast cells (MCs) are derived from hematopoietic stem cells and play important roles in allergic responses, innate immunity and defense against parasite infection. Unlike other blood cells, MCs migrate into peripheral tissues as immature progenitors and differentiate into mature mast cells. One of the unique features of MCs is that they show a variety of phenotypes depending on the different tissue microenvironment of their maturation [1]. In MCs, various MC-specific serine proteases are stored in the secretory granules, and their gene and protein expressions are dramatically altered when their cell environment is altered. For example, Reynolds *et al.* have shown that at least six distinct members of mouse MC-specific serine proteases are expressed in different combinations in different mast cell populations [2]. In addition, recent studies have shown that mature MCs vary in terms of what surface receptors and lipid mediators they express [3,4]. Because each mast cell population *in vivo* must play a specific role in the body, it is important to determine the character of each population of MCs.

Comprehensive gene expression analysis is a powerful approach to understand the characterization of various MC subpopulations. To date, several studies on microarray analysis of MCs have been conducted [5-7], but most of them dealt with MCs cultured *in vitro*. Alternatively, gene expression profiles of MCs isolated from skin and lung have been analyzed [3,8-10]. However, the numbers of MCs analyzed as one sample were relatively high and they were exposed to physical forces, enzymes and the anti-Kit antibody for purification, during which the original properties of the MCs may have been affected.

In the gastrointestinal tract, there are MCs that are mainly classified into two subclasses; mucosal MCs (mMCs) and submucosal MCs (sMCs) on the basis of their location, morphology (size and shape) and granule contents [11,12]. mMCs are mainly found in the mucosa of the gastrointestinal system, having chondroitin sulfate-containing granules, which are stained with toluidine blue but not safranin, and their activation occurs during parasite infection [13], while sMCs are localized in the submucosa of the gastrointestinal tract and their granules are rich in heparin and stained with both toluidine blue and safranin [1,11]. However, the molecular basis determining the differences in biochemical properties of these two MC subclasses remains uncertain, partially due to the difficulty of their isolation.

To overcome these problems, here we established a method of RNA amplification from intact MCs isolated from frozen tissue sections, which enables us to conveniently obtain the global gene expression pattern of MCs in various tissues. To validate this method, we first determined the minimum cell number required to achieve reproducible RNA amplification. We then compared the gene expression profiles obtained from small numbers of mMCs and sMCs in the mouse stomach, and found several key genes to be specifically expressed in one subclass of MCs, which may reflect some aspects of the distinct properties between the two MC subclasses in the gastrointestinal tract.

Results and discussion

Development of an RNA amplification protocol to obtain gene expression profiles from a small amount of RNA

To gain insight into the functional differences between the different subclasses of MCs, we employed three rounds of the T7-based RNA amplification method. Based on the preliminary experiments using peritoneal MCs and bone marrow-derived MCs (BMMCs), we estimated that a single MC yields 2 pg of RNA. Before we performed comparative analysis of MCs from different tissues, we first evaluated the accuracy and reproducibility of three rounds of the T7-based RNA amplification method, starting with the amount of RNA that can be obtained from a single MC. To assess this, we first compared the microarray results obtained from 5 μg of BMMC RNA prepared by the standard protocol with those obtained from the same RNA diluted 10^5- or 10^6-fold (30 pg, 10 pg and 2 pg) and subjected to three rounds of T7-based amplification (Figure 1a–c). Although three rounds of amplification yielded enough quantity of RNA for microarray analysis (>20 μg) even in the case of 2 pg RNA, scatter plot analysis revealed that the qualities of the obtained results were quite different between the samples from 5 μg and 2 pg RNA. The genes judged as 'Presence' in both 30 pg and 5 μg of RNA were 8,149 genes, which corresponded to 72% of genes judged as 'Presence' in the 5 μg of RNA (11,344 genes; Figure 1a), while only 4,116 genes were judged as 'Presence' in both 2 pg and 5 μg of RNA, which corresponded to only 36% of genes judged as 'Presence' in the 5 μg RNA (Figure 1c). The decrease in the number of genes judged as 'Presence' in the diluted samples (30 pg, 10 pg and 2 pg) may be due to the loss of low copy number RNA species during amplification.

We next examined the reproducibility of the microarray results obtained from two sets of 30 pg BMMC RNA samples (30 pg-1 and 30 pg-2) or two sets of 2 pg samples (2 pg-1 and 2 pg-2) (Figure 1d and 1e). In the 30 pg RNA samples, 7,537 (30 pg-1) and 8,777 (30 pg-2) genes were judged as 'Presence'. However, only 4,324 (2 pg-1) and 4,460 (2 pg-2) genes were judged as 'Presence' in each 2 pg RNA sample, again suggesting the loss of low copy number RNAs during amplification from a small amount of RNA. As to the reproducibility, 86% of the 'Presence' genes in the 30 pg-1 and 74% of 'Presence' genes in the 30 pg-2 sample were judged as 'Presence' in both 30 pg RNA

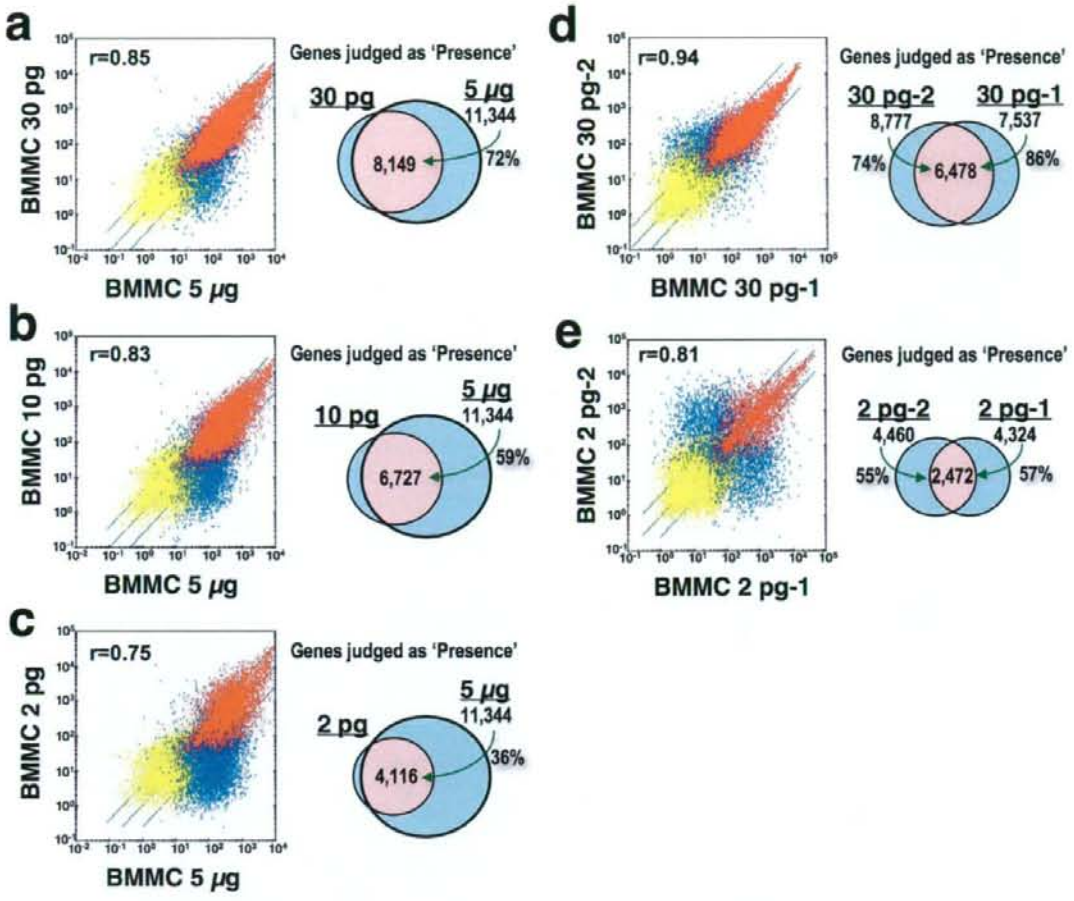


Figure 1

Comparisons of three round-amplified products starting with very small quantities of RNA. (**a-c**) Amplification biases in the products starting from a small quantity of RNA. Scatter plots of signal intensity obtained from 5 μg of BMMC RNA prepared by the standard protocol and from 30 pg (*a*), 10 pg (*b*) and 2 pg (*c*) of BMMC RNA prepared by three rounds of amplification are shown. (**d, e**) Reproducibility of the three-round amplification of a small quantity of RNA. Scatter plots of signal intensity between two independent products from 30 pg of BMMC RNA (BMMC 30 pg-1 and BMMC 30 pg-2) (*d*) or from 2 pg of BMMC RNA (BMMC 2 pg-1 and BMMC 2 pg-2) (*e*), are shown. Red dots show probe sets judged as "Presence", and yellow dots represent probe sets judged as "Absence" in both arrays. Blue dots show probe sets judged as "Presence" only in either array. The correlation coefficients (r) are presented. The same, four-fold induction and suppression thresholds are indicated as diagonal lines. Genes judged as "Presence" are placed in groups corresponding to pairwise overlaps shown in the accompanying Venn diagrams.

samples, while only 57% of 'Presence' genes in the 2 pg-1 and 55% of 'Presence' genes in the 2 pg-2 sample were judged as 'Presence' in both 2 pg RNA samples. These results suggested that the amplified products from the RNA from a single MC (about 2 pg) by the current method may include considerable amplification artifacts causing problems in accuracy and reproducibility. On the other hand, because of the higher reproducibility (>74%), we concluded that amplification from 30 pg RNA collected from 15 MCs would be suitable for the practical analysis of tissue MCs. Based on these results, we set our goal in this study to acquire gene expression profiles of MCs

pooled from different regions. To minimize the influence of cell-to-cell variations within the same class and potential amplification artifacts, we prepared three sets of 15 MCs for each region and compared genes with significantly different expression between MCs from the different regions (Figure 2b). We chose stomach as the source organ, since we can isolate two kinds of MCs from the mucosa (mMC) and the submucosa (sMC) regions of the same sections, and mMCs and sMCs have been suspected to be different in several MC properties such as protease expression profile and sensitivity to safranin staining [1,11].

Gene expression profiles of submucosal and mucosal MCs from the stomach

To visualize two kinds of MCs in the stomach without causing RNA degradation, the sections were fixed with carnoy's fixative and metachromatically stained with toluidine blue for a few seconds. sMCs and mMCs were microdissected using a patch pipette (Figure 2a and 2b). We prepared three sets of 15 MCs for each region, extracted their RNA and individually amplified them (sMC_1, sMC_2, sMC_3, and mMC_1, mMC_2, mMC_3). To improve the recovery of the extraction of as little as 30 pg of RNA, we used 'poly G' as a carrier, which does not interfere with the following RNA amplification or hybridization of the amplified product to the array (data not shown). To examine the effects of nonspecifically amplified artifact products, we performed the RNA extraction/amplification procedure without adding microdissected cells ("no cell") as a negative control (described in "*Materials and methods*"). The amplified RNAs of sMCs, mMCs and the "no cell" control were separately hybridized to a murine microarray. The signal values in the "no cell" sample were low in general and similar to the background levels (Figure 2c). The scatter plots of the samples independently prepared within the same group (e.g. sMC_1 vs sMC_2) showed a similar expression pattern; the average correlation coefficient for all probe-sets was 0.945 ± 0.004 and 0.893 ± 0.019 in sMCs and mMCs, respectively ($n = 3$). In contrast, the average correlation coefficient between sMCs and mMCs was 0.752 ± 0.034 ($n = 3$), which was much lower than those within the same group, suggesting that their gene expression patterns are different.

We further evaluated the accuracy and reproducibility of our method by other comprehensive analyses (hierarchical clustering analysis and principal component analysis [PCA]) using all probe sets. Microarray data obtained from sMCs, mMCs, skin-derived MCs, peritoneal MCs, BMMCs and non-MCs (macrophages and fibroblasts) were applied to these analyses. We first checked whether the amplification process in our method affects the global expression profile due to non-linear amplification. The results from the BMMC samples using RNA prepared by the standard protocol (BMMC-std) or the amplification method (BMMC-amp) were subjected to these analyses. Both hierarchical clustering analysis and PCA revealed that microarray data from BMMC-std and BMMC-amp were clustered in the same group (Figure 3a and 3b), suggesting that the global similarity in gene expression profiles is maintained during the amplification process. We next examined the similarity of expression patterns in three independent sMC or mMC samples. Upon clustering analysis and PCA, sMC_{1-3} and mMC_{1-3} were clustered in the same group, respectively. PCA also showed that the expression profiles of sMCs, mMCs and BMMCs are mutually different (Figure 3b).

We then compared the stomach-derived MCs (sMCs and mMCs) with skin-derived MCs, peritoneal MCs, BMMCs and non-MCs (macrophages and fibroblasts) by clustering analysis. The tissue-derived MCs (stomach MCs and skin MCs) were clustered separately from peritoneal MCs and BMMCs. These results may reflect different properties between tissue-derived MCs with firm adhesion to the neighboring cells and floating MCs without a tight contact. As to the similarity of MCs with fibroblasts and macrophages, it is reasonable that fibroblasts are most distant from MCs and macrophages are closer to MCs as a leukocyte family.

Validation of microarray results by real time RT-PCR analysis

We next investigated whether the hybridization signals of known marker genes specific for sMCs and mMCs showed the expected expression trends [12,14]. The mMC-specific genes, mast cell protease 1 (*Mcpt1*) and 2 (*Mcpt2*) showed higher values in mMCs, while the sMC-specific marker genes, mast cell protease 4 (*Mcpt4*) and chymase 2 (*Cma2*), showed higher signal values in sMCs (Table 1 and Figure 4a) [15-29]. On the other hand, MC-common markers such as kit oncogene (*Kit*) and Fcε receptor (*Fcer1a*) showed significant signal values with no bias between mMCs and sMCs. To further evaluate the results, we measured the expression levels of these marker genes by real-time RT-PCR using RNA from the independently isolated MCs (Figure 4b). Moreover, we randomly selected three genes showing 'mMC-biased' expression and another three genes showing 'sMC-biased' expression; expression of these genes in MCs has not been reported previously (Figure 4a). There were no significant differences in the expression levels of *Kit* and *Fcer1a* between mMCs and sMCs. In contrast, the mMC-specific markers *Mcpt1* and *Mcpt2* and the 'mMC-biased' genes, *Anxa10*, *Ctse*, and *Fos* showed higher expression in mMCs, and the sMC-specific markers *Mcpt4* and *Cma2* and the 'sMC-biased' genes, *Cnn1*, *Ces3*, and *Cpe* showed higher expression in sMCs. These results indicate that the microarray

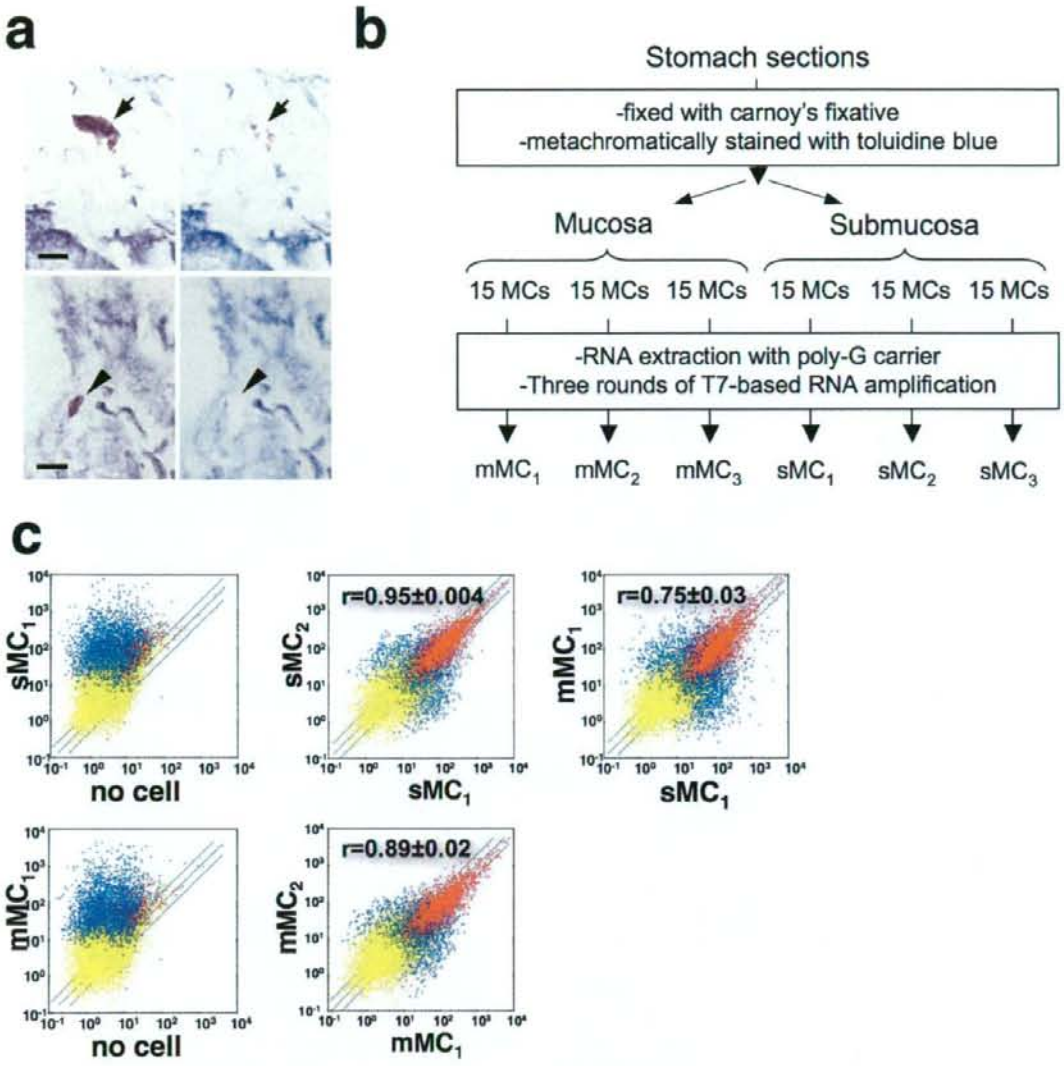


Figure 2

Gene expression profiles of sMCs and mMCs from stomach tissue. (**a**) Isolation of toluidine blue-stained MCs in the submucosa (sMC; *upper panels*) and the mucosa (mMC; *lower panels*) of stomach sections. A sMC (*arrow*) and mMC (*arrowhead*) that was metachromatically stained with toluidine blue before microdissection (*left panels*) disappeared after microdissection with a patch pipette (*right panels*). *Bars*, 10 μm. (**b**) Outline of the experimental strategy. (**c**) Labeled and fragmented antisense RNAs of three individual sMC samples, three individual mMC samples and the 'no cell' samples were hybridized to a Murine Array. Scatter plots for 'no cell' (x axis) and sMC_1 (y axis) (*upper left*), 'no cell' (x axis) and mMC_1 (y axis) (*lower left*), sMC_1 (x axis) and sMC_2 (y axis) (*upper center*), mMC_1 (x axis) and mMC_2 (y axis) (*lower center*), sMC_1 (x axis) and mMC_1 (y axis) (*upper right*) are shown. The correlation coefficients (r) for comparison within sMC_{1-3}, within mMC_{1-3} and between sMCs and mMCs are presented as means ± S.D. Red dots show probe sets judged as "Presence", and yellow dots represent probe sets judged as "Absence" in both arrays. Blue dots show probe sets judged as "Presence" only in either array. The same, two-fold induction and suppression thresholds are indicated as diagonal lines.

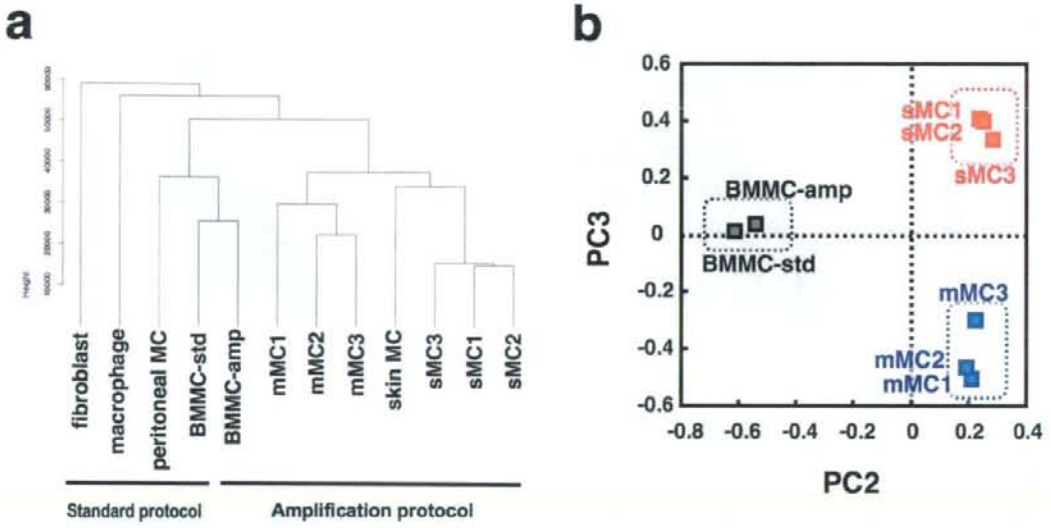


Figure 3
Global gene expression analysis of sMC_{1-3} and mMC_{1-3}. (**a**) Hierarchical clustering of global gene expression of various preparations of MCs and non-MCs. Three-round amplified products of sMC_{1-3}, mMC_{1-3}, skin MCs and BMMCs, and the standard products of BMMCs, peritoneal MCs, macrophages and fibroblasts were analyzed. (**b**) The principal component analysis (PCA) reveals different gene expression profiles of sMC_{1-3}, mMC_{1-3}, and two preparations of BMMCs. The blue dotted square indicates mMCs, the red dotted square indicates sMCs, and the black dotted square indicates BMMCs.

results are reliable and reflect the gene expression profiles of intact sMCs and mMCs in the stomach.

Clustering analysis of the gene expression profiles and functional categorization between sMCs and mMCs

Of the ~12,000 genes represented on the oligonucleotide array, we selected 1,272 genes whose expression levels between sMC_{1-3} and mMC_{1-3} were significantly different ($p < 0.05$, Limma's t test). The expression level of each gene was normalized by its level in BMMCs, which are cultured MCs with so-called 'immature' properties, and the selected genes were classified into seven clusters using the k-means clustering algorithm (CL1-7; Figure 5a and Additional file 1). We also classified the genes into functional categories, and the representative genes are listed (Figure 5b). Among them, 666 genes (52.4%) showed sMC-biased expression (*CL1*-3); in 78% (519 genes) of sMC-rich genes, the expression levels were relatively low in BMMCs and augmented in sMC (*CL1&2*). For example, the expression level of *Mcpt4* was relatively low in BMMCs, and if the expression profile of BMMCs reflects the immature properties of MC progenitors, *Mcpt4* can be concluded to be induced during the final maturation into sMCs. Interestingly, the sMC marker genes *Mcpt5* and *Mcpt6* were classified into *CL2/3*, suggesting that these genes were expressed to some extent in 'immature' BMMCs, but their expression was suppressed during maturation into mMCs. On the other hand, 606 genes (47.6%) showed mMC-biased expression (*CL4-7*); in 51% (334 genes) of mMC-rich genes, their expression levels in BMMCs were low but were augmented in mMCs (*CL4&5*). For example, expression of *Mcpt1* was low in 'immature' BMMCs but was drastically induced during maturation into mMCs.

Protein expression of Notch4 in sMCs and Ptgr1 in mMCs in stomach tissue

Among the genes showing differential expression (Figure 5b), we further focused on the expression of *Notch4* in sMCs and *Ptgr1* in mMCs, both of which have never been previously characterized in MCs. The *Notch4* gene product is a member of the Notch family, consisting of transmembrane receptors which are activated by cell surface ligands on adjacent cells. Recent studies have suggested that Notch signaling is involved in lymphocyte and mast cell differentiation [30,31]. We first confirmed that *Notch4* expression is significantly higher in the separately pooled sMCs than mMCs by real-time RT-PCR (data not shown).

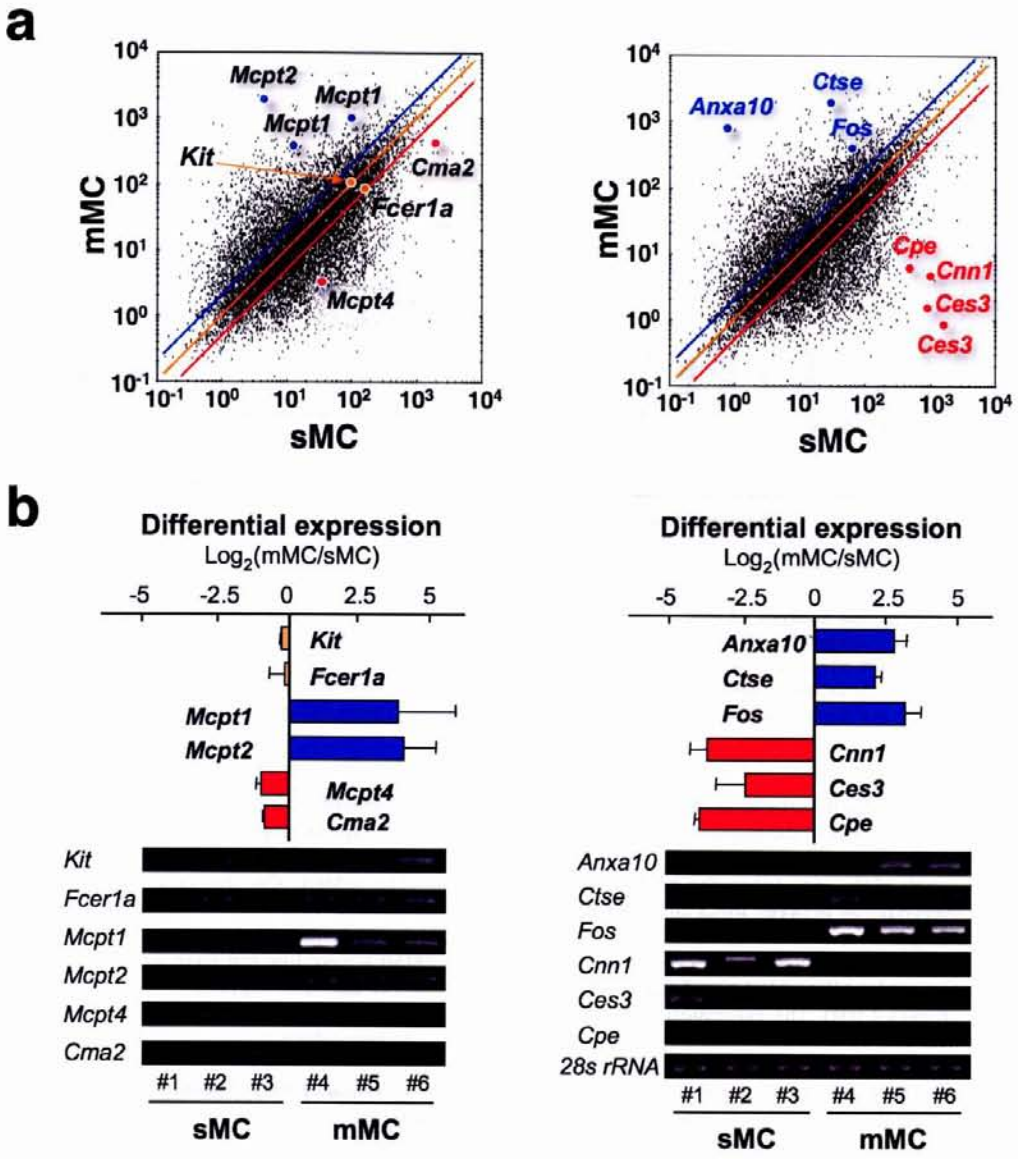


Figure 4

Validation of the differentially expressed genes between sMCs and mMCs. (**a**) sMC-specific (*Cma2*, *Mcpt4*), mMC-specific (*Mcpt1*, *Mcpt2*) and MC-common markers (*Fcer1a* and *Kit*) (*left panel*) and six randomly selected genes (*Ces3*, *Cnn1*, *Cpe*, *Anxa10*, *Ctse* and *Fos*) (*right panel*) are indicated in the representative scatter correlation graphs between sMC_I and mMC_I. The same, two-fold induction and suppression thresholds are indicated as a yellow, blue and red line, respectively. (**b**) The expression levels of the genes in (a) were verified by real-time RT-PCR. The values represent the ratio of relative expression levels of mMCs to sMCs, and are shown as mean ± S.D. (n = 3). The specificity of the PCR product was confirmed by gel electrophoresis and analysis of the melting temperature. The expression level of each gene was normalized to 28S ribosomal RNA.

Table 1: Summary of genes examined by real-time PCR analysis.

Gene Symbol	Gene Name	RefSeq Transcript ID	Reference
Kit	kit oncogene	NM_021099	15
Fcer1a	Fc fragment of IgE, high affinity I, receptor for α polypeptide	NM_010184	16
Mcpt1	mast cell protease 1	NM_008570	17, 18
Mcpt2	mast cell protease 2	NM_008571	19
Mcpt4	mast cell protease 4	NM_010779	2, 20
Cma2	chymase 2, mast cell (mast cell protease 10)	NM_001024714	14*
Anxa10	annexin A10	NM_011922	21
Ctse	cathepsin E	NM_007799	22
Fos	FBJ osteosarcoma oncogene	NM_010234	23
Ptgr1	Prostaglandin reductase 1 (leukotriene B_4 12-hydroxydehydrogenase)	NM_025968	24 (porcine)
Cnn1	calponin 1	NM_009922	25
Ces3	carboxylesterase 3	NM_053200	26
Cpe	carboxypeptidase E	NM_013494	27 (bovine)
Notch4	Notch gene homolog 4	NM_010929	28
28S rRNA	28S ribosomal RNA	NR_003279	29

*, The coding sequence presented in this paper is the N-terminus truncated-form of *Cma2*, while the RefSeq "NM_001024714" is the complete sequence of *Cma2*.

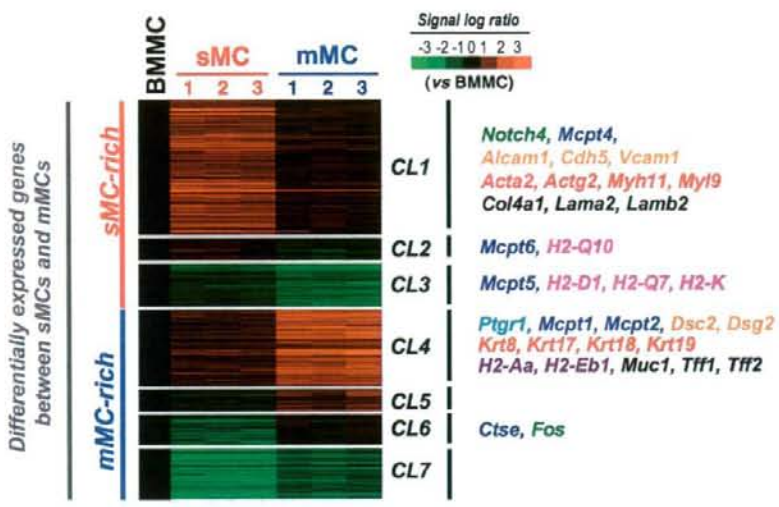
We next investigated whether the Notch4 protein is exclusively present in sMCs by immunostaining of stomach tissue (Figure 6a). Notch4 signals were detected in the nucleus-like structures of sMCs but not in those of mMCs. Furthermore, Notch4 signals were also found in the skin MCs, which were adjacently clustered with sMCs (Figure 3a). These results show that Notch4 is present in sMCs but not in mMCs, and suggest that Notch4 participates in sMC-specific transcription of Notch-target genes, which may be required for some sMC functions. In hematopoietic cells, it has been reported that constitutively active Notch4 promotes the expansion of progenitor cells and inhibits myeloid differentiation [32]. Since Notch ligands have been shown to exist in connective tissues such as skin dermis [33], it will be interesting to explore whether Notch4 plays a role in the differentiation of sMCs and the maintenance of sMC functions.

The *Ptgr1* product, 15-oxo-prostaglandin 13-reductase/leukotriene (LT) B_4 12-hydroxydehydrogenase is an essential enzyme for inactivation of eicosanoids such as prostaglandin E_2 (PGE_2) and LTB_4 [34]. Although it has been reported that the pathways of eicosanoid synthesis differ among the different MC subclasses [1,4], our results suggest that the inactivation system of eicosanoids also varies among the MC subclasses. *Ptgr1* expression was found to be significantly higher in the separately pooled mMCs by real-time RT-PCR (data not shown). We also examined Ptgr1 expression in stomach sections by immunostaining. Signals for the Ptgr1 protein were found in granule-like structures of mMCs in the stomach mucosa but not in sMCs (Figure 6b), suggesting that the Ptgr1 enzyme may be released from mMCs upon degranulation. Since PGE_2 plays critical roles in the maintenance of gut homeostasis through mucosal protection and inhibition of acid secretion, it is possible that when activated, mMCs negatively regulate the cytoprotective actions of PGE_2 through rapid inactivation by Ptgr1.

Gene expression pattern of extracellular matrix components, adhesion molecules, and cytoskeletal proteins in sMCs and mMCs

MC phenotypes have been shown to depend on their interactions with the surrounding extracellular matrices (ECMs) and neighboring cells [1]. One of the most remarkable findings in this study is the difference in gene expression of ECM protein components, adhesion molecules, and cytoskeletal proteins, which may reflect functional adaptation of each type of MC to the mucosal or submucosal environment in the stomach (Figure 5b). mMCs express genes for mucosa-specific ECM proteins such as *Muc1* (Mucin) and *Tff1* (Trefoil factor), while sMCs express genes for conventional ECM proteins such as *Col4a* (procollagen) and *Lama2* (laminin). Moreover, sMCs express genes for adhesion molecules such as *Alcam* and *Vcam1*, and genes for ordinary cytoskeletal proteins such as *Acta2* (actin), while mMCs express desmosome-component genes such as *Dsc2* (desmocollin) and *Dsg2* (desmoglein), and genes for keratin intermediate filaments such as *Krt8* and *Krt19*. Desmosomes were reported to be present in the stomach epithelia [35], and it was found that desmosome-like structures are detected in a particular type of MC [36]. It is thus possible that mMCs interact with adjacent epithelia through desmosomal adhesion in the stomach. In contrast, sMCs appear to interact with neighboring cells via adhesion molecules such as VCAM-1, ALCAM and VE-cadherin (*Vcam1*, *Alcam1* and *Cdh5*). Since these adhesion molecules have been shown to be involved in dynamic regulation of the actin cytoskeleton [37,38], such molecule-mediated inter-

a



b

Functional Category	sMC-rich (*CL1-3*)	mMC-rich (*CL4-7*)
MHC Class I	*H2-D1, H2-K, H2-Q7, H2-Q10*	
MHC Class II		*H2-Aa, H2-Eb1*
Protease	*Mpt4, Mcpt5, Mcpt6, Cma2, Cpe, Ctsl*	*Mcpt1, Mcpt2, Ctse*
Eicosanoid-related		*Ptgr1, Pla2g12*
Transcription-Associated Factor	*Notch4, Notch1, Barx1, Bcl6, Etv1, Madh1, Sox18*	*Foxa1, Fos, Sox2*
Signal Transduction	*Gnai1, Gnai2, Rgs2, Gnao, Sfrp1*	
Adhesion Molecule	*Alcam, Cdh5, Vcam1*	*Dsc2, Dsg2*
Receptor-related	*C1qr1, Eng, Il6st, Nptxr, Pdgfra,*	*Tnfrsf1b*
Cytoskeleton	*Acta2, Actg2, Cnn1, Des, Dstn, Myh11, Myl9, Pdlim3, Prph1, Stmn2, Tagln*	*Krt 8, Krt17, Krt18, Krt19*
Secretion		*Muc1, Tff1, Tff2*
Extracellular Matrix	*Bgn, Col4a1, Dcn, Fbln2, Lama2, Lamb2, Lum, Mglap, Pmp22, Sparcl1*	
Miscellaneous	*Ces3, Hdc*	*Anxa10*

Figure 5

Clustering analysis of the gene expression profiles between sMCs and mMCs. (**a**) Representation of mRNA expression levels of sMC_{1-3} and mMC_{1-3} compared with BMMCs. The color of the bars represents the ratio of signal intensity between independent samples and BMMCs, according to the scale shown on the *top right*. Genes with significantly different expression between sMCs and mMCs ($p < 0.05$, Limma's *t* test) were selected (1,272 genes) and classified into 7 clusters using the *k*-means algorithm (*CL1-7*). (**b**) Functional categorization of representative genes from (a).

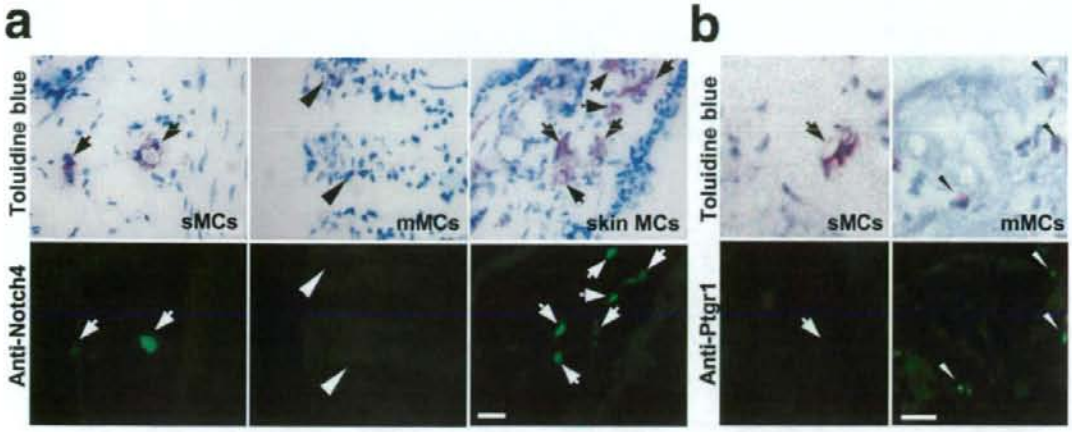


Figure 6
Immunohistochemical analysis of Notch4 and Ptgr1 in sMCs and mMCs in stomach tissue. (**a**) Stomach submucosa (sMCs; *left panels*), stomach mucosa (mMCs; *middle panels*) and skin (skin MCs; *right panels*) sections were stained with an anti-Notch4 antibody (*lower panels*) and with toluidine blue (*upper panels*). sMCs stained with the anti-Notch4 antibody in the gastric submucosa and skin dermis are indicated by arrows. No staining was observed in mMCs (*arrowheads*) localized in the gastric mucosa. sMCs and mMCs were metachromatically stained with toluidine blue. (**b**) Stomach submucosa (sMCs; *left panels*) and stomach mucosa (mMCs; *right panels*) sections were stained with an anti-Ptgr1 antibody (*lower panels*) and with toluidine blue (*upper panels*). No staining with the anti-Ptgr1 antibody was found in the sMCs (*arrow*). Small signals were observed in the mMCs (*arrowheads*). sMCs and mMCs were metachromatically stained with toluidine blue. *Bars*, 25 μm (a, b).

actions with submucosal cells may be critical to maintain the functional and morphological properties of sMCs. Indeed, it should be noted that most sMCs are variable in shape, and are often stretched and winding as compared with mMCs [1].

Conclusion

We established a method of RNA amplification from pooled intact MCs isolated from frozen tissue sections, which enables us to conveniently obtain the global gene expression pattern of MCs from various tissues, organs, and species including humans. By using this method, we demonstrated for the first time the distinct gene expression profiles of submucosal and mucosal MCs in the mouse stomach. Our findings offer insight into possible unidentified properties specific for each MC subclass.

Methods

Materials

The following materials were obtained from the sources indicated: HPLC purified T7-$(dT)_{24}$ primer [5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGC GG$(T)_{24}$] from GE Healthcare UK Ltd. (Buckinghamshire, England), RNase-free water, dNTP, SusperScript II, *Escherichia coli* (*E. coli*) RNase H, *E. coli* DNA polymerase I, *E. coli* DNA ligase, T4 DNA polymerase and random hexamers from Invitrogen (San Diego, CA), RNase inhibitor, glycogen, and MEGAscript T7 kit from Ambion (Austin, TX). Balb/c mice were obtained from JapanClea (Hamamatsu, Japan). This study was approved by the Committee on Animal Research of Kyoto University Graduate School of Pharmaceutical Sciences.

RNA amplification and oligonucleotide microarray

Mouse interleukin-3-dependent BMMCs were prepared as described previously [39]. Total RNA of BMMCs was extracted using RNeasy mini kit (Qiagen, Valencia, CA). Five micrograms of total RNA from BMMCs were labeled and prepared for hybridization according to the manufacturer's instructions (standard protocol). On the other hand, 30 pg, 10 pg and 2 pg of BMMC total RNA were amplified and labeled by our original three-round amplification method, which is described below.

First round

Total RNA was incubated with T7-$(dT)_{24}$ primer and first-strand cDNA was then synthesized by SuperScript II (Invitrogen). Second-strand synthesis was carried out by adding RNase H, DNA polymerase I and DNA ligase. The antisense RNA was synthesized using MEGAscript T7 kit.

Second round

The antisense RNA product was annealed with random hexamers, and cDNA was again synthesized by Super-

Script II. Then, the RNA-cDNA hybrid was digested with RNase H and annealed with the T7-$(dT)_{24}$ primer, and then second-strand synthesis was carried out by adding DNA polymerase I. The antisense RNA was again synthesized using MEGAscript T7 kit. Quality and size distribution of the antisense RNA product were confirmed by an RNA 6000 Nano LabChip on the Agilent Bioanalyzer (Palo Alto, CA).

Third round

As in the case of the second round, the double-stranded cDNA with a T7-promoter sequence was prepared from the second-round RNA product. Biotin-labeled antisense RNA was synthesized by RNA Transcript Labeling Kit (Enzo, Farmingdale, NY).

These labeled RNAs were hybridized to GeneChip Murine Expression oligonucleotide arrays (Affymetrix, Santa Clara, CA). We used microarray suite 5.0 of Affymetrix GeneChip Operating Software for quantification of the GeneChip data and decision of "Presence" or "Absence" of expression of each probe set using the values of 11 paired (perfect-matched and mismatched) probes.

Microdissection of MCs from tissue sections, RNA extraction, and microarray data analysis

Tissue sections 7 μm in thickness were prepared using a Jung Frigocut 3000E cryostat (Leica, Nussloch, Germany), and thaw-mounted onto poly-L-lysine-coated glass slides. To visualize MCs, the sections were fixed with carnoy's fixative, and immersed in toluidine blue using the following protocol: carnoy's fixative for 1 min, RNase-free water for 10 sec, toluidine blue (0.5% in 0.12N hydrochloric acid) for 5 sec, RNase-free water for 10 sec, 70% ethanol for 15 sec, and 100% ethanol for 15 sec three times; the sections were then vacuumed for 10 min to dry. Each single MC was microdissected from the sections using a patch pipette, and 15 cells were collected with an LCM Cap using the PixCell Ile Laser Capture Microdissection System (Arcturus, Mountain View, CA). As a negative control, LCM Caps just put on tissue sections without MCs were subjected to the same protocols (no cell). Fifteen microdissected MCs were homogenized in denaturing buffer of RNeasy mini kits. Twenty nanograms of poly G (Sigma, Saint Louis, MO) was added to the lysate as a nucleic acid carrier, and total RNA was extracted. Fifty picograms of BMMC total RNA (BMMC-amp) and total RNAs extracted from sMCs in the stomach submucosa, mMCs in the stomach mucosa and skin MCs in the ear dermis were amplified and labeled using the three-round amplification method, and were hybridized to U74Av2 Murine Genome Array (Affymetrix). On the other hand, total RNA of BMMCs (BMMC-std) and peritoneal MC, which were collected from mouse peritoneal cavities and purified by density gradient centrifugation using metrizamide, were labeled and hybridized by the standard protocol. Raw microarray data of macrophages (E-MEXP-38/298290452) and fibroblasts (E-GEOD-6697/1629511747) using the standard protocol were obtained from ArrayExpress, a public repository for transcriptomics data. We used either microarray suite 5.0 of Affymetrix GeneChip Operating Software or the robust multi-array average (RMA) expression measure for log transformation ($\log_2$) and normalization of the GeneChip data [40,41]. To determine the similarity in the data, hierarchical clustering analysis and PCA using the R statistical environment http://www.r-project.org were performed as a visualization technique. For comparison of the expression profiles of sMCs with that of mMCs, we selected 1,272 genes identified as having significantly different expression levels by the Limma's t-test ($p < 0.05$, $n = 3$). Signal values of sMCs and mMCs were normalized by the signal values of BMMCs. Using the k-means clustering algorithm, these genes were classified into seven clusters on the basis of similarity of their expression profiles.

Real-time reverse-transcription polymerase chain reaction (RT-PCR)

Total RNA extracted from 60 captured MCs was subjected to real-time RT-PCR. Real-time PCR was performed in a LightCycler (Roche, Mannheim, Germany) using Fast Start DNA Master SYBR Green I. The expression level of each gene was quantified using external standardized dilution, and normalized by 28S ribosomal RNA. Primer sequences are shown in Table 2. The specificity of the primers was confirmed by checking the product size and restriction enzyme pattern by gel electrophoresis and the melting temperature (data not shown).

Immunostaining

For tissue staining, frozen sections were fixed in 4% formaldehyde and incubated with a rabbit anti-Notch4 antibody (1:20, Santa Cruz Biotechnology, Santa Cruz, CA) or a rabbit anti-Ptgr1 antibody (1:20) which was a kind gift from Prof. Takao Shimizu (University of Tokyo) [42].

Abbreviations

BMMC: bone marrow-derived mast cell; CL: cluster; sMC: submucosa mast cell; DEPC: diethylpyrocarbonate; ECM: extracellular matrix; LCM: laser capture microdissection; LT: leukotriene; MC: mast cell; mMC: mucosa mast cell; PCA: principal component analysis; PG: prostaglandin; r: correlation coefficient; RMA: robust multi-array average; rRNA: ribosomal RNA; RT-PCR: reverse transcription-polymerase chain reaction.

Authors' contributions

ST designed the research, performed the research and wrote the paper; YT performed the research and wrote the paper; ES-N wrote the paper; YO performed the microar-

Table 2: List of primers used for real-time PCR analysis.

Gene Symbol	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
Kit	ATAGACCCGACGCAAC	AATAAACGAGTCACGCT
Fcer1a	GCCCCGTCTCCATTAG	CAATAACCCCGTGTCC
Mcpt1	AAACAGTCATAAATGGCAAG	GGGAACAAACCATCATCAC
Mcpt2	TTCATTGCCTAGTTCCTCT	CTTTTCAGCTACTTGCTCT
Mcpt4	CCTTACATGGCCCATCT	CTTCCCCGGCTTGATA
Cma2	GCGGAAATGCAAAGCC	ACAGGGAACAGTCCATC
Anxa10	TACCCACAACTTCGGC	GGCAAGTAGTGCTTTCT
Ctse	GCAAGCCTATTGGCAG	TGGCATCGTGTCGAGA
Fos	TGTGTACTCCCGTGGT	ACGAACAGGTAAGGTCC
Ptgr1	CATCGTGAATCGGTGG	GCTAGGTCAAACGCAT
Cnn1	ACGGCCTACGGTACAC	GGTACTCCGGGTTCAG
Ces3	AGTGATTGTGTCTCGAAG	GTTCCCATTCCGAGCA
Cpe	ACCGGAAGAGACTCTCA	CCAGTAATCCCCATCCT
Notch4	CCCTTAAACTCGGTTGT	GGTGCTTAATAAATAGTTGCC
28S rRNA	CAGTACGAATACAGACCG	GGCAACAACACATCATCAG

ray data and statistical analysis; ST performed the research; GT designed the research; ST performed the research; YS designed the research and wrote the paper. Conflict-of-interest: The authors declare no competing financial interests.

Additional material

Additional file 1

Genes with significantly different expression between sMCs and mMCs. The list represents 1,272 genes significantly altered between sMCs and mMCs in the order of clustering (Figure 5a). The values represent expression levels normalized to those of BMMCs.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-35-S1.xls]

Acknowledgements

This work was supported by Grants-in-Aid for Scientific Research on Priority Areas "Applied Genomics" from the Ministry of Education, Science, Sports, and Culture of Japan and from the Ministry of Health and Labor of Japan. We thank Dr. K Nakayama (Kyoto University) for their invaluable advice on this study. We appreciate Drs. T Shimizu and T Yokomizo for providing an anti-Ptgr1 antibody and generous instructions. We also thank Dr. HA Popiel and Ms. Y Nakaminami for careful reading and secretary assistance, respectively.

References

1. Metcalfe DD, Baram D, Mekori YA: **Mast cells.** *Physiol Rev* 1997, **77:**1033-1079.
2. Reynolds DS, Stevens RL, Lane WS, Carr MH, Austen KF, Serafin WE: **Different mouse mast cell populations express various combinations of at least six distinct mast cell serine proteases.** *Proc Natl Acad Sci USA* 1990, **87:**3230-3234.
3. Bradding P, Okayama Y, Kambe N, Saito H: **Ion channel gene expression in human lung, skin, and cord blood-derived mast cells.** *J Leukoc Biol* 2003, **73:**614-620.
4. Ogasawara T, Murakami M, Suzuki-Nishimura T, Uchida MK, Kudo I: **Mouse bone marrow-derived mast cells undergo exocytosis, prostanoid generation, and cytokine expression in response to G protein-activating polybasic compounds after coculture with fibroblasts in the presence of c-*kit* ligand.** *J Immunol* 1997, **158:**393-404.
5. Hernandez-Hansen V, Bard JD, Tarleton CA, Wilder JA, Lowell CA, Wilson BS, Oliver JM: **Increased expression of genes linked to FcepsilonRI Signaling and to cytokine and chemokine production in Lyn-deficient mast cells.** *J Immunol* 2005, **175:**7880-7888.
6. Iida M, Matsumoto K, Tomita H, Nakajima T, Akasawa A, Ohtani NY, Yoshida NL, Matsui K, Nakada A, Sugita Y, Shimizu Y, Wakahara S, Nakao T, Fujii Y, Ra C, Saito H: **Selective down-regulation of high-affinity IgE receptor (FcepsilonRI) alpha-chain messenger RNA among transcriptome in cord blood-derived versus adult peripheral blood-derived cultured human mast cells.** *Blood* 2001, **97:**1016-1022.
7. Nakajima T, Matsumoto K, Suto H, Tanaka K, Ebisawa M, Tomita H, Yuki K, Katsunuma T, Akasawa A, Hashida R, Sugita Y, Ogawa H, Ra C, Saito H: **Gene expression screening of human mast cells and eosinophils using high-density oligonucleotide probe arrays: abundant expression of major basic protein in mast cells.** *Blood* 2001, **98:**1127-1134.
8. Babina M, Schulke Y, Kirchhof L, Guhl S, Franke R, Bohm S, Zuberbier T, Henz BM, Gombart AF: **The transcription factor profile of human mast cells in comparison with monocytes and granulocytes.** *Cell Mol Life Sci* 2005, **62:**214-226.
9. Inomata N, Tomita H, Ikezawa Z, Saito H: **Differential gene expression profile between cord blood progenitor-derived and adult progenitor-derived human mast cells.** *Immunol Lett* 2005, **98:**265-271.
10. Okumura S, Kashiwakura J, Tomita H, Matsumoto K, Nakajima T, Saito H, Okayama Y: **Identification of specific gene expression profiles in human mast cells mediated by Toll-like receptor 4 and FcepsilonRI.** *Blood* 2003, **102:**2547-2554.
11. Enerback L: **Mast cells in rat gastrointestinal mucosa. 2. Dye-binding and metachromatic properties.** *Acta Pathol Microbiol Scand* 1966, **66:**303-312.
12. Friend DS, Ghildyal N, Austen KF, Gurish MF, Matsumoto R, Stevens RL: **Mast cells that reside at different locations in the jejunum of mice infected with Trichinella spiralis exhibit sequential changes in their granule ultrastructure and chymase phenotype.** *J Cell Biol* 1996, **135:**279-290.
13. Knight PA, Wright SH, Lawrence CE, Paterson YY, Miller HR: **Delayed expulsion of the nematode Trichinella spiralis in mice lacking the mucosal mast cell-specific granule chymase, mouse mast cell protease-1.** *J Exp Med* 2000, **192:**1849-1856.
14. Chu W, Johnson DA, Musich PR: **Molecular cloning and characterization of mouse mast cell chymases.** *Biochim Biophys Acta* 1992, **1121:**83-87.
15. Qiu FH, Ray P, Brown K, Barker PE, Jhanwar S, Ruddle FH, Besmer P: **Primary structure of c-kit: relationship with the CSF-1/**

PDGF receptor kinase family – oncogenic activation of v-kit involves deletion of extracellular domain and C terminus. *EMBO J* 1988, **7:**1003-1011.
16. Ra C, Jouvin MH, Kinet JP: **Complete structure of the mouse mast cell receptor for IgE (Fc epsilon RI) and surface expression of chimeric receptors (rat-mouse-human) on transfected cells.** *J Biol Chem* 1989, **264:**15323-15327.
17. Ghildyal N, McNeil HP, Stechschulte S, Austen KF, Silberstein D, Gurish MF, Somerville LL, Stevens RL: **IL-10 induces transcription of the gene for mouse mast cell protease-1, a serine protease preferentially expressed in mucosal mast cells of Trichinella spiralis-infected mice.** *J Immunol* 1992, **149:**2123-2129.
18. Trong HL, Newlands GF, Miller HR, Charbonneau H, Neurath H, Woodbury RG: **Amino acid sequence of a mouse mucosal mast cell protease.** *Biochemistry* 1989, **28:**391-395.
19. Serafin WE, Reynolds DS, Rogelj S, Lane WS, Conder GA, Johnson SS, Austen KF, Stevens RL: **Identification and molecular cloning of a novel mouse mucosal mast cell serine protease.** *J Biol Chem* 1990, **265:**423-429.
20. Serafin WE, Sullivan TP, Conder GA, Ebrahimi A, Marcham P, Johnson SS, Austen KF, Reynolds DS: **Cloning of the cDNA and gene for mouse mast cell protease 4. Demonstration of its late transcription in mast cell subclasses and analysis of its homology to subclass-specific neutral proteases of the mouse and rat.** *J Biol Chem* 1991, **266:**1934-1941.
21. Morgan RO, Jenkins NA, Gilbert DJ, Copeland NG, Balsara BR, Testa JR, Fernandez MP: **Novel human and mouse annexin A10 are linked to the genome duplications during early chordate evolution.** *Genomics* 1999, **60:**40-49.
22. Tatnell PJ, Lees WE, Kay J: **Cloning, expression and characterisation of murine procathepsin E.** *FEBS Lett* 1997, **408:**62-66.
23. Curran T, MacConnell WP, van Straaten F, Verma IM: **Structure of the FBJ murine osteosarcoma virus genome: molecular cloning of its associated helper virus and the cellular homolog of the v-fos gene from mouse and human cells.** *Mol Cell Biol* 1983, **3:**914-921.
24. Yokomizo T, Izumi T, Takahashi T, Kasama T, Kobayashi Y, Sato F, Taketani Y, Shimizu T: **Enzymatic inactivation of leukotriene B4 by a novel enzyme found in the porcine kidney. Purification and properties of leukotriene B4 12-hydroxydehydrogenase.** *J Biol Chem* 1993, **268:**18128-18135.
25. Strasser P, Gimona M, Moessler H, Herzog M, Small JV: **Mammalian calponin. Identification and expression of genetic variants.** *FEBS Lett* 1993, **330:**13-18.
26. Ovnic M, Tepperman K, Medda S, Elliott RW, Stephenson DA, Grant SG, Ganschow RE: **Characterization of a murine cDNA encoding a member of the carboxylesterase multigene family.** *Genomics* 1991, **9:**344-354.
27. Fricker LD, Evans CJ, Esch FS, Herbert E: **Cloning and sequence analysis of cDNA for bovine carboxypeptidase E.** *Nature* 1986, **323:**461-464.
28. Uyttendaele H, Marazzi G, Wu G, Yan Q, Sassoon D, Kitajewski J: **Notch4/int-3, a mammary proto-oncogene, is an endothelial cell-specific mammalian Notch gene.** *Development* 1996, **122:**2251-2259.
29. Hassouna N, Michot B, Bachellerie JP: **The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes.** *Nucleic Acids Res* 1984, **12:**3563-3583.
30. Radtke F, Wilson A, Mancini SJ, MacDonald HR: **Notch regulation of lymphocyte development and function.** *Nat Immunol* 2004, **5:**247-253.
31. Sakata-Yanagimoto M, Nakagami-Yamaguchi E, Saito T, Kumano K, Yasutomo K, Ogawa S, Kurokawa M, Chiba S: **Coordinate regulation of transcription factors through *Notch2* is an important mediator of mast cell fate.** *Proc Natl Acad Sci USA* 2008, **22:**7839-7844.
32. Vercauteren SM, Sutherland HJ: **Constitutively active Notch4 promotes early human hematopoietic progenitor cell maintenance while inhibiting differentiation and causes lymphoid abnormalities in vivo.** *Blood* 2004, **104:**2315-2322.
33. Nickoloff BJ, Qin JZ, Chaturvedi V, Denning MF, Bonish B, Miele L: **Jagged-1 mediated activation of notch signaling induces complete maturation of human keratinocytes through NF-kappaB and PPARgamma.** *Cell Death Differ* 2002, **9:**842-855.
34. Hori T, Yokomizo T, Ago H, Sugahara M, Ueno G, Yamamoto M, Kumasaka T, Shimizu T, Miyano M: **Structural basis of leukotriene B4 12-hydroxydehydrogenase/15-Oxo-prostaglandin 13-reductase catalytic mechanism and a possible Src homology 3 domain binding loop.** *J Biol Chem* 2004, **279:**22615-22623.
35. Brennan D, Hu Y, Kljuic A, Choi Y, Joubeh S, Bashkin M, Wahl J, Fertala A, Pulkkinen L, Uitto J, Christiano AM, Panteleyev A, Mahoney MG: **Differential structural properties and expression patterns suggest functional significance for multiple mouse desmoglein 1 isoforms.** *Differentiation* 2004, **72:**434-449.
36. Vodenicharov A, Chouchkov C: **Morphological study of mast cell localization in the wall of the proximal tubule in the domestic swine kidney.** *Anat Histol Embryol* 1999, **28:**85-88.
37. Swart GW, Lunter PC, Kilsdonk JW, Kempen LC: **Activated leukocyte cell adhesion molecule (ALCAM/CD166): signaling at the divide of melanoma cell clustering and cell migration?** *Cancer Metastasis Rev* 2005, **24:**223-236.
38. Cook-Mills JM: **VCAM-1 signals during lymphocyte migration: role of reactive oxygen species.** *Mol Immunol* 2002, **39:**499-508.
39. Tanaka S, Takasu Y, Mikura S, Satoh N, Ichikawa A: **Antigen-independent induction of histamine synthesis by immunoglobulin E in mouse bone marrow-derived mast cells.** *J Exp Med* 2002, **196:**229-235.
40. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data.** *Nucleic Acids Res* 2003, **31:**e15.
41. Sugimoto Y, Tsuboi H, Okuno Y, Tamba S, Tsuchiya S, Tsujimoto G, Ichikawa A: **Microarray evaluation of EP4 receptor-mediated prostaglandin E2 suppression of 3T3-L1 adipocyte differentiation.** *Biochem Biophys Res Commun* 2004, **322:**911-917.
42. Yamamoto T, Yokomizo T, Nakao A, Izumi T, Shimizu T: **Immunohistochemical localization of guinea-pig leukotriene B4 12-hydroxydehydrogenase/15-keto prostaglandin 13-reductase.** *Eur J Biochem* 2001, **268:**6105-6113.

最近の話題

日薬理誌 133, 173 (2009)

ケミカルゲノミクスに基づくインシリコ創薬

新島　聡，奥野　恭史

長年，創薬においては，副作用を抑えるべく，単一のタンパク質に選択的に作用して機能を制御する低分子化合物（リガンド）の設計が追求されてきた．しかし一方では，単一の化合物がファミリーの全く異なる複数のタンパク質と作用する例や，単一のタンパク質が多様な化合物と結合する例など，多対多の化合物－タンパク質間相互作用が数多く報告されている(1)．実際に，がん，精神疾患，感染症の治療薬としてすでに有用性が認められている医薬品の中にも，想定外のタンパク質に作用するものが少なくないことが明らかになってきた(2)．最近では，これら多分子標的作用に基づく薬理学が 'Polypharmacology'（多重薬理学）と称され注目されつつある(3)．単一の標的タンパク質に作用する薬剤を開発する場合，'polypharmacology' は好ましくない性質と見なされるが，一分子対一分子の関係に着目したこれまでのアプローチでは，複合性の疾患には対処できないなどの限界があり，むしろ 'polypharmacology' を利用して，複数のタンパク質を標的とするような薬剤を開発する必要性が高まっているのも現状である．

「ケミカルゲノミクス」とは，ゲノムプロジェクトによって集積された大量のバイオ情報を化合物と関連づけ，遺伝子・タンパク質と膨大な数の化合物との相互作用を包括的に明らかにすることを命題とした新興分野であり，マイクロアレイやハイスループットスクリーニングなど，近年の技術革新を背景に，膨大な化合物に関する情報収集が行われ，有用な化合物の探索が加速度的に展開されている．これらの大規模データは，多対多の化合物－タンパク質間相互作用関係を潜在的に網羅していると考えられ，'polypharmacology' を理解する上で，極めて重要な知見をもたらすことが期待される(4)．

著者らは，膨大なケミカルゲノミクスデータから，化合物群とタンパク質群との相互作用関係を効率的に抽出し，実用的なインシリコ創薬につなげることを主眼において，データベースの開発および，体系的な相互作用予測法の研究開発を進めてきた．具体的には，創薬標的として最も重要なファミリーの一つである G タンパク質共役型受容体（GPCR）を対象として，相互作用情報を収集・整備し，データベース GLIDA (GPCR-LIgand DAtabase) として Web 上で公開している(5)．さらには，情報科学技術の一つである機械学習を用いて，化合物とタンパク質との結合情報（ケミカルゲノミクス情報）からタンパク質のリガンド認識パターンを抽出することにより，活性化合物を効率的に探索する，'Chemical Genomics-based virtual screening (CGBVS)' という新規なバーチャル・スクリーニング（VS）手法を提唱している．VS の従来手法としては，既知リガンドとの構造類似性に基づく 'ligand-based virtual screening (LBVS)' と，標的タンパク質の立体構造に基づく 'structure-based virtual screening (SBVS)'（ドッキング計算）がよく知られているが，CGBVS は LBVS とは異なり，リガンド未知のタンパク質，すなわちオーファンタンパク質に対する相互作用予測をも可能にし，一方 SBVS と比較して，計算効率性が非常に高いという利点を有する．また，CGBVS は必ずしもタンパク質の立体構造情報を必要としないため，結晶構造を得ることが困難な膜タンパク質に対するリガンド予測に適している．

実際に，CGBVS と LBVS の性能比較を行うために，GPCR を対象に大規模な計算機実験を行い，ケミカルゲノミクス情報の活用が，相互作用予測の性能向上につながることを実証している．さらに，X 線結晶構造が最近明らかにされたヒト β_2 アドレナリン受容体に焦点をあて，CGBVS の予測スコア（活性化合物である確率を表す）が上位 50 個の化合物について in vitro 結合阻害実験を行ったところ，実に全体では 89% (31/35)，μM オーダでは 14% という驚異的なヒット率を示した（図 1）．また，LBVS では難しいとされる，新規骨格を持つリガンドの検出にも有効であることが示唆された．新規化合物の探索は，新薬開発の特許戦略上において重要であることから，この結果は CGBVS の実用性の高さを示すものといえる．次に，X 線結晶構造を用いたドッキング計算による予測と比較したところ，CGBVS はドッキングよりも高い予測率を示した．特に計算コストにおいては，通常のドッキン

キーワード：ケミカルゲノミクス，化合物 - タンパク質間相互作用，バーチャル・スクリーニング
京都大学大学院薬学研究科 寄附講座システム創薬科学（〒606-8501 京都市左京区吉田下阿達町 46-29）
E-mail: okuno@pharm.kyoto-u.ac.jp　Satoshi Niijima, Yasushi Okuno
原稿受領日：2008 年 10 月 16 日