AIGR Deep Research Report: MBL2 and Surfactant Homeostasis (GO:0043129)

Gene: MBL2 (mannose-binding lectin 2) · **Organism:** *Homo sapiens* (NCBITaxon:9606) · **UniProt:** P11226 (Mannose-binding protein C) **Focus type:** function_assignment · **Hypothesis slug:** function-hypothesis-go-0043129 **Seed hypothesis:** *MBL2 has surfactant homeostasis (GO:0043129).* **Annotation under review:** GO:0043129 "surfactant homeostasis", evidence **IBA** (ECO:0000318), reference **GO_REF:0000033**

Summary

The proposed function assignment — that MBL2 directly participates in **surfactant homeostasis (GO:0043129)** — is **over-annotated and refuted as a direct function of MBL2.** On MBL2 (P11226) the term exists **only** as an IBA (phylogenetic) annotation from GO_REF:0000033, with **zero** experimental support (no IDA, IMP, IPI, or IGI) anywhere in MBL2's ~94 GO annotations. The annotation is best explained as a **paralog carry-over** propagated across the collectin family tree from the true pulmonary surfactant collectins, **SP-D (SFTPD)** and SP-A (SFTPA1/2).

The biology of MBL2 is inconsistent with an alveolar surfactant role. MBL2 is a **liver-synthesized, secreted serum collectin** whose experimentally characterized function is initiation of the **lectin pathway of complement** — Ca^{2+}-dependent recognition of microbial carbohydrate arrays, MASP protease activation, and downstream opsonization/phagocytosis. UniProt annotates it as "secreted" and "produced mainly in the liver," with no alveolar/lung localization and no surfactant lipid- or protein-handling activity. Surfactant homeostasis, by contrast, is an alveolar process executed by SP-A/SP-D, the lamellar-body lipid transporter ABCA3, and surfactant-processing proteases (CTSH, NAPSA) — genes that carry GO:0043129 with genuine experimental (IDA/IMP/ISS) support.

The one structural fact that both explains the mis-annotation and cautions the curator is that MBL2 and SP-D are close **collectin paralogs** — both built from an N-terminal collagen-like region plus a C-type lectin carbohydrate-recognition domain (CRD), with ~56% full-length identity and ~50% CRD identity. This kinship is exactly what allows a phylogenetic annotation

pipeline to project a lung-branch function onto a serum-branch member. The similarity is real; the functional inference is not. **Recommended curation lead (requires curator verification): remove or down-rank GO:0043129 on MBL2 as a non-core, low-confidence IBA carry-over, retaining the term on SFTPD/SFTPA where it is experimentally supported.**

Key Findings

Finding F001 — GO:0043129 on MBL2 is an unsupported IBA paralog carry-over from SP-D

The single confirmed finding of this investigation is that **MBL2's surfactant-homeostasis annotation has no gene-specific experimental basis and is a computational inheritance from its surfactant-collectin paralogs.** The evidence has four interlocking parts.

1. Evidence provenance: the term is IBA-only. In the QuickGO annotation record for MBL2 (P11226), GO:0043129 is carried **solely** as IBA (ECO:0000318, GO_REF:0000033). Across MBL2's full complement of ~94 GO annotations there is **no** experimental evidence code attached to the surfactant term. IBA ("Inferred from Biological Ancestor") is produced by the GO Phylogenetic Annotation (PAINT/PAN-GO) project: a curator annotates ancestral nodes of a gene tree, and the annotation is propagated to descendant genes. It is a *prediction based on family membership*, and by GO convention it is explicitly non-experimental. IBA is a procedurally valid evidence code — it is not "wrong" as a mechanism — but here it represents biological over-reach.

2. Where the experimental support actually lives. Of the 16 human genes annotated to GO:0043129, experimental support is concentrated in genes with direct, measured surfactant biology, whereas MBL2 sits in an IBA-only tier alongside an adhesion GPCR and a pseudogene:

Gene	Role	Best evidence for GO:0043129
SFTPD (SP-D)	Pulmonary surfactant collectin	IDA (PMID 19265061), IMP (PMID 9751757)
BPIFA1	Airway/surfactant-associated	IDA (PMID 23499554)
CTSH / NAPSA	Surfactant protein processing proteases	IDA (PMID 18216060)
ABCA3	Lamellar-body lipid transporter	ISS
MBL2	Serum lectin (complement)	**IBA only**
ADGRF5	Adhesion GPCR	IBA only
BPIFA4P	Pseudogene	IBA only

The company MBL2 keeps in the IBA-only tier is itself a red flag: genuine surfactant biology comes with experimental annotations; MBL2's does not.

3. Structural paralogy drives the carry-over. MBL2 is a **collectin** — signal peptide (1–20), an N-terminal cysteine-rich/collagen-like region (~42–99), a coiled-coil neck (~112–130), and a C-terminal **C-type lectin CRD** (~134–245). SP-D (SFTPD) shares this exact modular architecture. Pairwise Needleman–Wunsch alignment gives full-length identity of **56.1%** and CRD identity of **~50%** between MBL2 and SFTPD (for comparison, MBL2–SFTPA1 ~35% and SFTPD–SFTPA1 ~51%). In a gene-family tree, MBL2 and SFTPD sit close together, so any term annotated to a surfactant-collectin ancestor can be projected onto MBL2 by IBA even though MBL2 never acquired surfactant function.

4. MBL2 biology contradicts an alveolar/surfactant role. UniProt (P11226) describes MBL2 as **secreted** and a **"plasma protein produced mainly in the liver."** Its annotated and experimentally supported function is a **Ca^{2+}-dependent lectin that initiates the lectin pathway of complement** — binding mannose/GlcNAc on microbial surfaces, oligomerizing into higher-order "bouquets," complexing with MASP proteases, and driving opsonization and phagocytosis. MBL2's experimentally supported GO terms include GO:0001867 (complement

activation, lectin pathway; IDA), P 9087411 pattern-recognition-receptor activity, mannose binding (GO:0005537), and antibacterial defense. There is **no** UniProt localization to lung, alveolar type II cells, or the surfactant lining layer, and **no** annotated participation in surfactant lipid/protein turnover.

Interpretation. GO:0043129 on MBL2 reflects **family-level inference, not gene-level function.** The term is legitimately experimental on SP-D/SP-A and legitimately propagates *within* the pulmonary-collectin subfamily, but MBL2 is a serum complement collectin in a different compartment performing a different process. The annotation is therefore an over-annotation on MBL2.

Mechanistic Model / Interpretation

The crux of this case is distinguishing **shared architecture** from **shared function**. Collectins are a family defined by a collagen-like stalk plus a C-type lectin CRD, but different collectins are deployed in different anatomical compartments for different biological jobs.

```
                    COLLECTIN FAMILY (collagen region + C-type lectin CRD)
                                        |
     ---------------------------------------------------------------------
     |                                  |                                  |
SERUM / COMPLEMENT                 PULMONARY / SURFACTANT            OTHER (CL-L1, CL-K1...)
     |                                  |
  MBL2 (P11226)                  SFTPD (SP-D), SFTPA1/2 (SP-A)
liver-derived, secreted          alveolar type II / Clara cells
     |                                  |
Ca2+-dependent binding to         Ca2+-dependent binding to surfactant
microbial sugar arrays            lipids/proteins + pathogens
     |                                  |
MASP-1/2/3 complexes              regulate surfactant pool size,
-> lectin complement pathway       structure, and reuptake
-> opsonization / phagocytosis    -> SURFACTANT HOMEOSTASIS (GO:0043129)
     |                                  |
[INNATE IMMUNITY core]            [SURFACTANT HOMEOSTASIS core]
     |
     └── GO:0043129 attaches to MBL2 ONLY by IBA (phylogenetic carry-over),
         NOT by any experimental measurement of MBL2.
```

Both branches share a Ca^{2+}-dependent CRD and a collagen stalk, so a phylogenetic annotation pipeline readily projects a term from the pulmonary branch onto the serum branch. But the **compartment** (serum vs. alveolus), the **cellular source** (hepatocyte vs. alveolar type II/Clara cell), and the **downstream effector system** (MASP/complement vs. surfactant lipid turnover) all differ. Surfactant homeostasis is a property of the SP-D/SP-A branch; its appearance on MBL2 is a false positive of family-level inference.

A useful mental test for the curator: *Would loss of MBL2 perturb alveolar surfactant pool size or composition?* The published MBL2 loss-of-function literature centers on **infection susceptibility, complement activity, and immune modulation** — not on pulmonary surfactant metabolism, and not on alveolar proteinosis (the SP-D knockout phenotype). MBL2 shares the molecular *tool* (a Ca^{2+}-dependent CRD) but not the biological *process*.

Evidence Base

The evidence matrix below summarizes the most decision-relevant items. A striking pattern emerges: **every experimentally grounded description of MBL2 function points to the lectin complement pathway and innate immunity, and none describes MBL2 participating in surfactant metabolism.** The experimental home of GO:0043129 is SP-D, not MBL2.

Evidence Matrix

#	Citation	Evidence type	Supports/ Refutes/ Qualifies/ Competing	Claim tested	Key finding	Context
1	QuickGO (P11226), GO_REF:0000033 (ECO:0000318)	Review/ database (provenance)	**Refutes** (as direct)	Is GO:0043129 on MBL2 experimentally supported?	GO:0043129 on MBL2 is **IBA-only**; 0 experimental codes among ~94 annotations	Human, database-
2	SFTPD: PMID:19265061 (IDA), PMID:9751757 (IMP)	Direct assay / mutant phenotype	Qualifies (localizes true function)	Which gene truly owns surfactant homeostasis?	Experimental GO:0043129 belongs to **SP-D**, not MBL2	Human/m lung
3	UniProt P11226 (curated)	Review/ database	**Refutes**	Is MBL2 a lung surfactant protein?	"Secreted"; "produced mainly in the liver"; function = Ca^{2+}-dependent lectin initiating lectin complement pathway	Human, plasma/liv
4	Sequence analysis (this run, NW alignment)	Structural/ evolutionary (computed)	Explains mechanism of error	Is MBL2 a paralog of SP-D?	MBL2–SFTPD 56.1% full-length, ~50% CRD; same collagen+C-type-lectin architecture	Human collectins
5	PMID:9087411	Direct assay	**Competing** (correct function)	What is MBL2's assayed activity?	MBL2 → complement activation, lectin pathway	Human se

#	Citation	Evidence type	Supports/ Refutes/ Qualifies/ Competing	Claim tested	Key finding	Context
					(GO:0001867) = innate-immunity core	
6	PMID:15060079	Direct assay / mechanism	Competing	MBL molecular mechanism	MBP (MBL) binds surface carbohydrates and activates MASP-1/2/3 → C4/C2 cleavage → complement	Serum biochemis
7	PMID:21091907	Mutant/ phenotype (human deficiency)	Competing	MBL2 loss-of-function phenotype	MBL deficiency alters dendritic-cell innate/antigen-presenting function (immune, not pulmonary)	Human bl
8	PMID:16410008, PMID:18952132	Genetic	Competing	MBL2 functional variation	Exon-1 SNPs → impaired polymerization, low serum MBL, ↑ infection susceptibility	Human populatio
9	PMID:32335925, PMID:19606686	Association	Competing	MBL2 physiological correlates	Low serum MBL associates with adverse pregnancy outcomes / gastritis via innate	Human se

#	Citation	Evidence type	Supports/ Refutes/ Qualifies/ Competing	Claim tested	Key finding	Context
					immunity/ opsonization	
10	PMID:22475410, PMID:23814060	Review / direct assay	Qualifies	Collectin family compartmentalization	Collectins share architecture but differ in tissue/role; CL-L1 binds mannose via CRD, forms MASP complexes → lectin pathway	Human plasma / review
11	PMID:10589993, PMID:10931846	Structural/ biochemical	Qualifies	CRD specificity divergence	Serum- vs. liver-type MBP CRDs differ in oligosaccharide affinity; CRD specificity is modular/ tunable	Rat MBPs, crystallog

How the literature bears on the hypothesis

Across the retrieved primary literature, **every experimentally grounded description of MBL2 function points to the lectin complement pathway and innate immunity** — carbohydrate recognition, MASP association, opsonization, phagocytosis, and infection/disease susceptibility (PMID:15060079, PMID:21091907, PMID:16410008, PMID:18952132). **None** describes MBL2 participating in alveolar surfactant metabolism. Family reviews (PMID:22475410) and CL-L1 work (PMID:23814060) reinforce that collectins share architecture but are **compartmentalized** into distinct functional niches — the structural basis for why an IBA pipeline mis-assigns the surfactant term to MBL2. The experimental home of GO:0043129 is SP-D (PMID:19265061, PMID:9751757), not MBL2.

GO Curation Implications

Term: GO:0043129 "surfactant homeostasis" (Biological Process). **Current action on MBL2:** annotated, IBA (ECO:0000318), GO_REF:0000033.

Recommended curation action (lead — requires curator verification):

- **Remove or down-rank GO:0043129 on MBL2 as a non-core, over-annotated term.** It is IBA-only, conflicts with MBL2's liver/serum localization, and its experimental basis in the family resides in the SP-A/SP-D lung branch. If retained for procedural completeness, explicitly flag it as a low-confidence phylogenetic carry-over that does not represent MBL2's biology.
- **Retain GO:0043129 on the genuine surfactant genes** where it is experimentally supported — SFTPD (IDA/IMP), plus BPIFA1, CTSH/NAPSA, ABCA3.
- **Elevate MBL2's experimentally supported terms** as its core annotations:
- **BP:** complement activation, lectin pathway (**GO:0001867**, IDA P 9087411); innate immune / antimicrobial signaling; defense response to Gram-positive bacterium (GO:0050830).
- **MF:** mannose binding (**GO:0005537**); pattern recognition receptor activity (GO:0038187); Ca^{2+}-dependent protein binding (GO:0048306).
- **CC:** extracellular region / secreted (GO:0005576). Note that MBL2's IBA-only CC term GO:0005771 (multivesicular body) also warrants curator scrutiny.
- **Do not** default to "protein binding" (GO:0005515); the specific mannose-binding-lectin / lectin-complement-pathway terms are supported and preferable.

MF / BP / CC judgment: GO:0043129 is a BP term. The evidence does **not** support this BP for MBL2. It supports a *different* BP (lectin complement pathway) plus specific MF/CC terms. The surfactant BP should be **removed/generalized-away** on MBL2, not retained as core.

Mechanistic Scope

The immediate molecular activity under interrogation is whether MBL2's **Ca^{2+}-dependent C-type lectin CRD** participates in **surfactant homeostasis** — regulation of the pool size, composition, or turnover of pulmonary surfactant lipids/proteins in the alveolar lining layer.

- **Direct gene-product activity of MBL2:** Ca^{2+}-dependent binding of terminal mannose/GlcNAc on microbial glycan arrays (CRD, ~134–245); oligomerization via the N-terminal collagen-like region (~42–99) and coiled-coil neck (~112–130) into higher-order bouquets; formation of complexes with MASP-1/2/3; triggering of the lectin complement pathway (C4/C2 cleavage) → opsonization and phagocytosis. This is the *measured* function.
- **Surfactant homeostasis (the tested term):** a distinct **alveolar** process executed by SP-A/SP-D (surfactant collectins), ABCA3 (lamellar-body lipid transport), and surfactant-processing proteases (CTSH, NAPSA). MBL2 has no measured activity in this process, no alveolar type-II-cell expression, and no lamellar-body role.
- **Downstream/pleiotropic effects of MBL2 loss** (increased infection susceptibility, altered dendritic-cell cytokines, disease associations) are **immune** phenotypes, not pulmonary surfactant phenotypes — they do not support GO:0043129.

The distinction matters because IBA can conflate a **shared molecular tool** (a Ca^{2+}-dependent CRD) with a **shared biological process** (surfactant homeostasis). MBL2 shares the tool but not the process.

Conflicts and Alternatives

1. **Paralog confusion (the dominant alternative, and the likely truth).** MBL2 and SP-D/SP-A are close collectin paralogs (~56% full-length identity, ~50% CRD identity). The surfactant term legitimately belongs to the pulmonary collectin branch and is projected onto MBL2 by phylogenetic inference. This is the parsimonious explanation for the IBA annotation.
2. **Compartment mismatch.** MBL2 is serum/liver-derived; surfactant homeostasis is alveolar. No primary literature places MBL2 in the alveolar lining fluid as a surfactant regulator.
3. **Database carry-over.** GO_REF:0000033 is the standard reference for GO phylogenetic (IBA) annotations, confirming the term's origin is the PAN-GO/PAINT pipeline rather than a curated experimental paper.

4. **Competing core function.** MBL2's experimentally supported role — lectin complement pathway / innate immunity (PMID:9087411, PMID:15060079) — is a strong, well-documented alternative that should anchor the review.
5. **No isoform or artifact rescue.** MBL2 has a single characterized secreted product; there is no lung-expressed isoform performing surfactant functions. The discrepancy is evolutionary annotation carry-over, not an isoform-specific role.

No retrieved evidence conflicts with the refutation; all alternatives point the same way (family carry-over, not a real MBL2 surfactant function).

Limitations and Knowledge Gaps

- **Database snapshot dependency.** The finding rests on the current QuickGO/GO annotation state (evidence codes, GO_REF). A curator should re-verify that GO:0043129 on MBL2 remains IBA-only at review time.
- **Negative-evidence caveat.** Absence of an experimental annotation is not proof of absence of function. **What was checked:** MBL2 GO provenance, the GO:0043129 gene set, UniProt localization/function, and paralogy. **The gap:** no direct experimental assay was found testing MBL2 for surfactant-lipid binding or alveolar surfactant turnover. **Why it matters:** the crux is IBA-only status; a targeted negative experiment (e.g., MBL2 CRD vs. DPPC binding) would formally close it. **Resolver:** a deep literature search for "mannose-binding lectin AND pulmonary surfactant / DPPC / alveolar" — none surfaced in the immunity-focused searches performed.
- **Ancestral-node question.** Whether the PAN-GO/PANTHER ancestral node annotation should itself be re-curated (which would also affect ADGRF5 and BPIFA4P IBA) is unresolved; resolver is a PAN-GO tree review.
- **Provenance re-verification.** The SFTPD experimental references (
  P 19265061
  IDA;
  P 9751757
  IMP) and MBL2's
  P 9087411

lectin-pathway support were taken from the annotation graph; their exact snippets should be pulled and confirmed before citation.

- **Alignment parameters.** The reported identities (56.1% full-length, ~50% CRD) came from pairwise NW alignment of MBL2 vs. SFTPD; a curator should confirm the exact isoform and parameters used.
- **Non-human mechanistic data.** Some CRD-specificity detail derives from rat MBP-A/MBP-C (PMID:10589993); human specifics may differ, though the compartmentalization principle holds.

Discriminating Tests

To decisively separate "MBL2 has surfactant homeostasis" from "MBL2 is a serum complement lectin mis-annotated by paralogy":

1. **Expression/localization check.** Query GTEx/Human Protein Atlas: confirm MBL2 mRNA/protein is liver-restricted with negligible alveolar type-II-cell expression, while SP-D is lung-dominant. (Programmatic, decisive.)
2. **Annotation provenance audit.** Confirm via QuickGO API that GO:0043129 on P11226 is ECO:0000318 (IBA)/GO_REF:0000033 with no IDA/IMP/IPI. (Directly resolves the curation question.)
3. **Binding assay.** Test MBL2 CRD for DPPC/surfactant-lipid binding vs. SP-D positive control (predict negative/nonspecific).
4. **Knockout phenotype comparison.** MBL2-deficient humans/mice show infection susceptibility, not alveolar proteinosis/surfactant accumulation (contrast with SP-D KO). Any surfactant phenotype would be discriminating.
5. **PAN-GO tree audit.** Inspect where in the collectin tree GO:0043129 was placed and whether the serum-collectin subclade should be pruned.

Proposed Follow-up Experiments / Actions

For the curator (immediate, low-cost): - Re-pull the QuickGO annotation for P11226 and confirm GO:0043129 is IBA-only; if so, **flag for removal/down-ranking** on MBL2. - Verify that the SFTPD references (P 19265061, P 9751757) support GO:0043129 on SFTPD, and retain the term there. - Retrieve P 9087411 and extract the exact snippet documenting MBL2's lectin-pathway experimental annotation before using it as MBL2's core BP support. - Also scrutinize MBL2's other IBA-only terms — GO:0005771 (multivesicular body, CC) and GO:0050766 (positive regulation of phagocytosis, the latter biologically plausible via opsonization and possibly upgradeable to experimental).

Computational provenance to run/record: - MBL2 vs. SFTPD global and CRD-only NW alignment, saving the alignment and percent-identity table. - MBL2 tissue-expression pull (GTEx/HPA), saving the expression table showing liver dominance. - PANTHER/PAINT node inspection to document the ancestral origin of the IBA propagation.

Curation Leads (require curator verification):

Lead	Detail
Action change	Remove or down-rank GO:0043129 (surfactant homeostasis, IBA, GO_REF:0000033) on MBL2 as an over-annotation.
Retain elsewhere	Keep GO:0043129 on SFTPD/SFTPA (experimentally supported IDA/IMP).
Candidate core terms for MBL2	BP: GO:0001867 (complement activation, lectin pathway), GO:0050830 (defense to Gram-positive bacterium); MF: GO:0005537 (mannose binding), GO:0038187 (PRR activity); CC: GO:0005576 (extracellular region, secreted).
Candidate references to verify (exact snippets)	MBL2 lectin-pathway IDA P 9087411; SFTPD surfactant homeostasis IDA P 19265061, IMP P 9751757.
Suggested question	"Is any MBL2 experimental paper reporting a pulmonary-surfactant function? If none, GO:0043129 should not be propagated to MBL2."
Suggested experiment	Comparative KO phenotype + tissue-expression audit (MBL2 vs. SFTPD) to confirm functional divergence.

Conclusion

The seed hypothesis that **MBL2 has surfactant homeostasis (GO:0043129)** is **over-annotated / refuted as a direct function.** The annotation is an **IBA (phylogenetic) paralog carry-over** with zero MBL2-specific experimental support, propagated from the true pulmonary surfactant collectins SP-D/SP-A across a structurally similar but functionally distinct family. MBL2's experimentally established role is initiating the **lectin pathway of complement** as a **liver-derived, secreted serum lectin** in innate immunity. The recommended lead is to remove or down-rank GO:0043129 on MBL2 while retaining it on the genuine surfactant genes, and to anchor the MBL2 review on its complement/innate-immunity core functions.