

LUNG AND COLON CANCER

Lung and Colon Cancer: Documentation

I. Benign Tissue Categories (Healthy Tissues)

These sections serve as a reference for the AI to identify normal tissues and establish the baseline for non-malignancy.

1. Colon Benign Tissue (Healthy Colon Tissues)

Category	Key Technical Data	Relevance for AI (Input/Output)
Normal Histology	Regular and aligned Crypts of Lieberkühn , formed by enterocytes (absorptive cells) and goblet cells (mucus-producing).	Input: Recognition of regular nuclear polarity and glandular architecture.
Cytology	Small, basal, oval, and uniform nuclei; abundant cytoplasm, lack of significant hyperchromasia or pleomorphism.	RAG Output: Low probability of mutations (e.g., <i>KRAS</i> or <i>BRAF</i>).
Immunomarkers	Expression of CK20 (pos) and CK7 (neg).	Immunophenotypic Baseline to rule out malignancy.

2. Lung Benign Tissue (Healthy Lung Tissues)

Category	Key Technical Data	Relevance for AI (Input/Output)
Normal Histology (Alveoli)	Thin, intact alveoli lined by Type I Pneumocytes (gas exchange) and Type II Pneumocytes (surfactant).	Input: Identification of non-compressed, non-invaded alveolar structure.
Normal Histology (Bronchi)	Ciliated pseudostratified columnar epithelium with goblet cells.	Input: Recognition of normal bronchial/bronchiolar architecture.
Immunomarkers	Type II Pneumocytes: positive for TTF-1 (Thyroid Transcription Factor-1) and Napsin A .	Immunophenotypic Baseline for lung tissue (useful in distinguishing primary from metastatic tumors).

II. Malignant Neoplasms

3. Colon Adenocarcinoma (Cancerous cells of the colon)

Category	Key Technical Data	Relevance for AI (Input/Output)
Histological Definition	Malignant proliferation of epithelial cells forming atypical glandular structures or producing mucus. Lack of maturation and loss of polarity.	Input: Detection of "dirty" glands (necrotic) and back-to-back arrangement (without intervening stroma).

Category	Key Technical Data	Relevance for AI (Input/Output)
Cytology	Elongated, pleomorphic, hyperchromatic nuclei with atypical mitoses; reduced or absent normal goblet cells.	RAG Output: Grade of differentiation (G1-G3).
Staging (TNM)	Penetration of the intestinal wall (<i>T</i>), lymph node involvement (<i>N</i>), distant metastases (<i>M</i>).	Input: T, N, M stage (determines surgical/oncological <i>management</i>).
Molecular Biology (Crucial for RAG)	Microsatellite Instability (MSI) or Mismatch Repair Deficiency (dMMR). <i>KRAS</i> , <i>NRAS</i> , <i>BRAF</i> V600E mutations.	RAG Output for Therapy: Anti-EGFR therapy (if <i>RAS</i> Wild Type), Immunotherapy (if MSI-H/dMMR).
Immunomarkers	CK20 (pos), CDX2 (pos), CK7 (neg).	Differential Diagnosis (vs. Lung Adenocarcinoma).
ICD-O Code	8140/3 (Adenocarcinoma, NOS).	Output: Coding and reporting.

4. Lung Adenocarcinoma (Cancerous cells of the lung)

Category	Key Technical Data	Relevance for AI (Input/Output)
Histological Definition	Malignant proliferation of epithelial cells with glandular differentiation or mucus production. Subtypes: Lepidic, Acinar, Papillary, Solid, Micropapillary.	Input: Classification of invasive patterns (more invasive = worse prognosis). RAG
Cytology	Large, pleomorphic nuclei, variable cytoplasm (sometimes mucinous); cells growing along existing alveolar walls (lepidic pattern).	Output: Definition of the primary <i>driver</i> (distinction from metastasis).
Staging (TNM)	Similar to Colon, but with emphasis on pleural/airspace invasion (T) and hilar/mediastinal involvement (N).	Input: Stage (guides surgery/radio-chemotherapy). RAG Output for Therapy: Targeted Therapies (TKIs).
Molecular Biology (Crucial for RAG)	Activating <i>driver</i> mutations: EGFR (exone 19 deletion, L858R), ALK , ROS1 rearrangements, BRAF V600E, MET .	The AI must prioritize searching for these targets.
Immunomarkers	TTF-1 (pos), Napsin A (pos), CK7 (pos), CK20 (neg).	Differential Diagnosis (vs. Metastatic Colon Adenocarcinoma).
ICD-O Code	8140/3 (Adenocarcinoma, NOS).	Output: Coding and reporting.

5. Lung Squamous Cell Carcinoma (Aggressive lung cancer type)

Category	Key Technical Data	Relevance for AI (Input/Output)
Histological Definition	Malignant proliferation showing squamous cell differentiation . Characterized by keratin pearls and intercellular bridges (desmosomes).	Input: Recognition of keratinization and polygonal cell morphology.
Cytology	Large, polygonal cells with abundant eosinophilic cytoplasm and hyperchromatic nuclei; tendency to form nests.	RAG Output: Correlation with smoking history (stronger than Adenocarcinoma).
Typical Location	Typically central (proximal to the main bronchi).	Input: CT (Computed Tomography) data for location (central vs. peripheral).
Molecular Biology	Less dependent on <i>driver</i> mutations than Adenocarcinoma. Often TP53 mutations and SOX2 amplification. PD-L1 expression is important for immunotherapy.	RAG Output for Therapy: Chemotherapy and/or Immunotherapy (anti-PD-1/PD-L1) .
Immunomarkers	p40 (pos) , CK5/6 (pos) , TTF-1 (neg), Napsin A (neg).	Differential Diagnosis (vs. Adenocarcinoma).
ICD-O Code	8070/3 (Squamous Cell Carcinoma, NOS).	Output: Coding and reporting.

Lung and Colon Cancer for Clinicians and AI/RAG Applications

This document covers the three most common histopathological entities in public AI datasets (e.g., LC25000, IHC colorectal datasets) and their benign counterparts:

- Colon adenocarcinoma (COAD)
- Lung adenocarcinoma (LUAD)
- Lung squamous cell carcinoma (LUSC)
- Benign colonic tissue
- Benign lung tissue

1. Colon Adenocarcinoma (COAD)

Epidemiology (2025 data): ~153,000 new cases/year USA; 10–15% hereditary (Lynch, FAP); median age 68 y; left-sided > right-sided in sporadic cases.

Molecular Pathogenesis (Vogelstein model + updates)

- Chromosomal instability (CIN) pathway (85%): APC → KRAS → TP53 → 18q LOH
- Microsatellite instability (MSI-H) pathway (15%): MLH1 hypermethylation or Lynch syndrome
- CpG island methylator phenotype (CIMP) Key drivers: BRAF V600E (right-sided, poor prog), HER2 amp (5–10%), NTRK fusions (rare).

Histopathology (WHO 5th ed. Digestive Tumors 2019/2025 update)

Feature	Description	Frequency
Glandular architecture	Tubular, cribriform, solid nests; mucinous (>50% extracellular mucin)	90%
Cytology	Columnar cells, nuclear stratification, hyperchromasia, prominent nucleoli	—
Grading (WHO)	G1 well-diff (>95% glandular), G2 moderate (50–95%), G3 poor (<50%)	G2 most common
Special subtypes	Mucinous, signet-ring, serrated, medullary (MSI-H), micropapillary (aggressive)	10–15%
Invasion	Desmoplastic stroma, tumor budding (≥ 10 buds = Bd3, poor prog), LVI	—

Staging (AJCC 8th ed. + 2025 updates)

- T4 subclassification: T4a (peritoneal penetration), T4b (organ invasion)
- Prognostic stage groups incorporate MSI status and molecular markers.

Prognosis & Biomarkers

- Stage I: >90% 5-y OS; Stage IV: ~15%
- MSI-H/dMMR: excellent prognosis with immunotherapy (pembrolizumab 1st line)
- RAS mutated: no benefit from EGFR inhibitors
- BRAF V600E: poor prognosis, targeted with encorafenib + cetuximab
- HER2+: trastuzumab-deruxtecan approved 2024–2025
- Consensus Molecular Subtypes (CMS1–4) for research.

2. Colon Benign Tissue

Normal colonic mucosa and benign lesions mistaken for cancer in AI datasets.

Entity	Histopathology Features	Key to Distinguish from Adenocarcinoma
Normal colonic mucosa	Orderly crypts, basal nuclei, goblet cells, Paneth cells	No dysplasia, regular architecture
Hyperplastic polyp	Serrated crypts, no dysplasia	Lack of cytologic atypia
Sessile serrated lesion (SSL)	Dilated crypts with basal dilatation, horizontal growth	No conventional dysplasia (but pre-malignant)
Tubular adenoma (low-grade dysplasia)	Crowded glands, pencil-like nuclei	Dysplasia confined to epithelium (no invasion)

3. Lung Adenocarcinoma (LUAD)

Epidemiology: Most common NSCLC (40–50%); increasing in never-smokers (especially women, EGFR mutations); strong association with smoking but 10–25% never-smokers.

Molecular Drivers (2025)

- EGFR mutations (ex19del, L858R): 15% US, 40–60% Asia
- KRAS (G12C most common): 25–30%
- ALK rearrangements: 4–7%
- ROS1, RET, NTRK, MET ex14, BRAF V600E, HER2
- STK11/LKB1 and KEAP1 co-mutations: poor prognosis and immunotherapy resistance
- ~70% of advanced LUAD are actionable.

Histopathology (WHO 5th ed. Thoracic Tumors 2021/2025 update)

Pattern (IASLC grading)	Description	Frequency	Prognostic Weight
Lepidic (AIS/MIA)	Growth along alveoli, non-invasive	10–20%	Best
Acinar	Gland formation	Most common	Intermediate
Papillary/Micropapillary	Papillae or small clusters without fibrovascular cores	10–15%	Worst
Solid	Sheets of cells, no recognizable pattern	—	Worst
Invasive mucinous	Mucin-filled alveoli, lepidic growth (former mucinous BAC)	5%	KRAS predominant

Grading (IASLC 2024)

- Grade 1: Lepidic predominant
- Grade 2: Acinar/papillary predominant
- Grade 3: Solid/micropapillary/complex glandular predominant

Spread Pattern Scoring (2025): % of each pattern; predominant + highest-grade pattern determine final grade.

4. Lung Squamous Cell Carcinoma (LUSC)

Epidemiology: 20–30% NSCLC; strongly smoking-related (>95%); central location; median age 70 y.

Molecular Features

- TP53 almost universal
- Actionable alterations rare: FGFR1 amp, PI3KCA, DDR2 (trials)
- High tumor mutational burden → good immunotherapy response
- PD-L1 expression often high (>50% in 30–50% cases)

Histopathology

Subtype/Feature	Description	Frequency
Keratinizing	Keratin pearls, intercellular bridges	60%
Non-keratinizing	No keratin but squamous differentiation by p40/p63	30%
Basaloid	High N/C ratio, peripheral palisading	5–10%

Subtype/Feature	Description	Frequency
Grading	G1 well (keratinizing), G3 poor (basaloid/solid)	G2–G3 most

Immunohistochemistry to confirm squamous: p40 (most specific), p63, CK5/6; TTF-1 negative.

5. Lung Benign Tissue

Entity	Histopathology Features	Key to Distinguish from Malignancy
Normal lung	Alveoli with type I/II pneumocytes, bronchioles, vessels	Orderly architecture, no atypia
Reactive pneumocytes	Type II hyperplasia (cuboidal, enlarged nuclei)	Preserved architecture, no invasion
Organizing pneumonia	Intra-alveolar fibroblasts, chronic inflammation	Polypoid plugs, no malignant cells
Granulomas	Epithelioid histiocytes, giant cells	Special stains negative for organisms

Key Immunohistochemistry Panel (for AI and Pathology Confirmation)

Marker	Colon Adeno	Colon Benign	Lung Adeno	Lung SCC	Lung Benign
CK20	+ (80–90%)	+	–	–	Variable
CK7	– (90%)	–	+ (95%)	– (90%)	+ (bronchial)
CDX2	+ (90%)	+	–	–	–
TTF-1	–	–	+ (90%)	–	+ (type II)
Napsin A	–	–	+	–	+ (type II)
p40	–	–	–	+ (95%)	–
Chromogranin/Synaptophysin	–	–	–	–	– (except neuroendocrine)

Staging Summary (AJCC 8th ed., applicable 2025)

Cancer Type	Key Prognostic Features Beyond Stage
Colon Adeno	MSI status, RAS/BRAF, sidedness, CMS, tumor budding, circumferential margin
Lung Adeno	Predominant + high-grade pattern (IASLC grade), spread-through-air-spaces (STAS), PD-L1, driver mutations
Lung SCC	PD-L1 expression (main predictive marker), TMB, location (central vs peripheral)

TNM Staging and Molecular Markers

I. TNM Staging (Tumor, Node, Metastasis)

The TNM stage, defined by the AJCC (American Joint Committee on Cancer), is the most critical prognostic system, as it dictates the treatment pathway (surgery, radiotherapy, chemotherapy, immunotherapy).

A. Lung Cancer (Adenocarcinoma and Squamous Cell Carcinoma)

Lung cancer staging is particularly complex due to the distinction between parenchymal invasion and airway/pleural involvement.

TNM Category	Technical Definition (Lung)	Prognostic/Therapeutic Implication for AI
T (Primary Tumor)	Based on size (T1a <1 cm, T4 ≥7 cm or involvement of vital structures like the diaphragm, heart, great vessels) and local invasion (e.g., invasion of the visceral pleura or tumor nodules in the same lobe).	T1-T2a: Often candidates for surgical resection (lobectomy). T4: Often requires multimodality therapy or is considered unresectable.
N (Regional Lymph Nodes)	Based on the location of the involved lymph nodes: N1 (Ipsilateral peribronchial or hilar), N2 (Ipsilateral mediastinal or subcarinal), N3 (Contralateral mediastinal/hilar or supraclavicular).	N0-N1: Better prognosis. N2: Often treated with chemoradiation before surgery or as definitive therapy. N3: Advanced disease, typically managed with systemic therapy.
M (Distant Metastasis)	M0 (absent), M1a (pulmonary nodules in the opposite lobe or malignant pleural/pericardial effusion), M1b/M1c (metastases to distant organs, e.g., brain, bones, liver).	M1: Metastatic disease (Stage IV). AI Focus: Systemic therapy (targeted drugs or immunotherapy) rather than curative surgery.

B. Colon Adenocarcinoma

TNM Category	Technical Definition (Colon)	Prognostic/Therapeutic Implication for AI
T (Primary Tumor)	Based on the depth of invasion of the intestinal wall : T1 (submucosa), T2 (muscularis propria), T3 (through the muscularis propria into pericolic/perirectal tissue), T4 (invasion of the serosa or adjacent structures).	T1-T2: Surgical resection. T3-T4: Often requires adjuvant chemotherapy post-surgery to reduce recurrence risk.
N (Regional Lymph Nodes)	Based on the number of positive regional lymph nodes: N1 (1–3 positive nodes), N2 (≥4 positive nodes).	N0: Best prognosis. N1-N2: Clear indication for adjuvant chemotherapy (e.g., FOLFOX). The AI must prioritize this recommendation.

TNM Category	Technical Definition (Colon)	Prognostic/Therapeutic Implication for AI
M (Distant Metastasis)	M0 (absent), M1a (metastasis limited to one organ), M1b (metastasis in more than one organ/peritoneal site).	M1: Metastatic disease. AI Focus: Personalized systemic therapy based on molecular status (<i>KRAS</i> , <i>BRAF</i> , MSI).

II. Prognostic Implications of Molecular Markers

Molecular analysis not only identifies therapeutic targets but also stratifies risk and potential drug response, which is crucial information for the RAG system.

A. Lung Cancer (Adenocarcinoma)

Molecular Marker	Prognostic/Therapeutic Implication	Value for AI/RAG
<i>EGFR</i> Mutations	Prognosis without Treatment: Usually worse prognosis. Treatment: Excellent response to first, second, or third-generation TKIs (Tyrosine Kinase Inhibitors) (e.g., Osimertinib).	RAG Output: If <i>EGFR</i> is mutated, the AI must propose TKI therapy as the standard of care.
<i>ALK</i> Rearrangements	Prognosis without Treatment: Poor prognosis. Treatment: Excellent and durable response to ALK Inhibitors (e.g., Alectinib).	RAG Output: If <i>ALK</i> is rearranged, the AI must propose specific ALK TKI therapy.
<i>PD-L1</i> (Expression)	Prognosis: High expression (TPS) is associated with a better response to immunotherapy. Treatment: Immunotherapy (anti-PD-1, e.g., Pembrolizumab) monotherapy for high PD-L1 expressing tumors.	RAG Output: The percentage of PD-L1 expression determines the choice between TKI, immunotherapy alone, or chemotherapy + immunotherapy.

B. Colon Adenocarcinoma

Molecular Marker	Prognostic/Therapeutic Implication	Value for AI/RAG
Mutated <i>KRAS</i>	Prognosis: Unfavorable prognosis. Treatment: Intrinsic resistance to anti-EGFR therapy (e.g., Cetuximab/Panitumumab).	RAG Output: Exclusion of anti-EGFR therapy. Indication for standard chemotherapy or anti-KRAS G12C drugs (if specific mutation).

Molecular Marker	Prognostic/Therapeutic Implication	Value for AI/RAG
Mutated <i>BRAF</i>V600E	Prognosis: Very unfavorable (shorter survival), often associated with MSS and aggressive metastatic presentation. Treatment: Requires combination therapy (e.g., Targeted Therapy + TKI) to overcome resistance.	RAG Output: Flags an aggressive subtype that requires intensive regimens.
MSI-H / dMMR (Microsatellite Instability / Mismatch Repair Deficiency)	Prognosis: Better prognosis in early stages. Treatment: Exceptional and durable response to Immunotherapy (Checkpoint Inhibitors).	RAG Output: For metastatic disease (M1), the AI should suggest immunotherapy as first-line treatment, regardless of the metastatic site.

Cross-cutting “anchor” resources

These give you almost everything a doctor needs for **colon and lung cancer**:

- **ESMO Clinical Practice Guidelines – Gastrointestinal Cancers (Colon):** incidence, risk factors, diagnostic work-up, staging, treatment algorithms for early and metastatic colorectal cancer. [ESMO+1](#)
- **SIGN / BMJ colorectal cancer guidelines:** practical, evidence-based recommendations spanning prevention, diagnosis, surgery, oncology, and follow-up. [sign.ac.uk+1](#)
- **ESMO Clinical Practice Guidelines – Lung and Chest Tumours / NSCLC:** full coverage of lung adenocarcinoma and lung squamous cell carcinoma (NSCLC) including staging and treatment. [ESMO+1](#)
- **NICE Lung Cancer NG122:** diagnostic pathways, imaging, bronchoscopy/biopsy, MDT, systemic therapy, and palliative care. [NICE](#)

For AI, these provide high-level clinical context, staging rules (TNM), and management standards.

2. Colon Adenocarcinoma (Cancerous cells of the colon)

Pathology / biology

- **UNSW “Gastrointestinal Tract – Colon Histology”:** contrasts normal colon with neoplastic changes, useful for benign vs adenocarcinoma. [embryology.med.unsw.edu.au](#)

- **Rook / Robbins chapters on colorectal carcinoma** (textbooks; if you have PDFs): adenoma–carcinoma sequence, molecular pathways (APC, KRAS, MMR), growth patterns, histologic grading.

Clinical & staging

- **ESMO “Early colon cancer” and “Metastatic colorectal cancer”**: diagnostic algorithm (colonoscopy + biopsy, CT, MRI), TNM staging and treatment. [ESMO+2Annals of Oncology+2](#)
-

3. Colon Benign Tissue (Healthy colon tissues)

You need high-quality **normal histology** references so the model can clearly distinguish benign tissue:

- **NUS Pathweb – Colon Normal Histology**: annotated virtual slides showing normal crypt architecture, goblet cells, mucosa/submucosa/muscularis. medicine.nus.edu.sg
- **Histology Guide – Colon (large intestine)**: detailed images and labels of normal colonic layers and mucosal glands. histologyguide.com
- **Libre Pathology / ScienceDirect topics – Normal colonic mucosa**: descriptive text on cell types, normal architecture, and how it differs from dysplasia and carcinoma. librepathology.org+2ScienceDirect+2

These are perfect to index as “benign colon” for your class **Colon Benign Tissue**.

4. Lung Adenocarcinoma (Cancerous cells of the lung)

Classification & pathology

- **2015 and 2021 WHO Classification of Lung Tumors summaries**: explain adenocarcinoma patterns (lepidic, acinar, papillary, micropapillary, solid) and new grading systems. jcd.net+3Jto+3ScienceDirect+3
- **Libre Pathology – Lung adenocarcinoma**: concise overview of morphology (usually peripheral, gland-forming tumors), immunohistochemistry (TTF-1, Napsin A) and molecular features (EGFR, ALK, ROS1). librepathology.org
- **ICCR “Lung cancer resection dataset”**: proforma for path reports—tumor type, grade, margins, lymphovascular invasion, etc. iccr-cancer.org

Clinical

- Use the **ESMO NSCLC** and **NICE lung cancer** guidelines above for staging (TNM), risk stratification, and management.
-

5. Lung Squamous Cell Carcinoma (Aggressive NSCLC subtype)

- **WHO / 2021 updates on lung squamous carcinoma:** details keratinizing, non-keratinizing, basaloid types and associated molecular aspects. [pathologyjournal.rcpa.edu.au+3pubs.rsna.org+3jcdt.net+3](#)
 - **Wikidoc – Squamous cell carcinoma of the lung classification:** quick reference to WHO subtypes, risk factors (strong smoking association), and basic staging. [Wikidoc](#)
 - **Diagnostic surgical pathology in lung cancer guidelines:** emphasizes why distinguishing adenocarcinoma vs squamous histology is critical for therapy choice. [Guideline Central](#)
-

6. Lung Benign Tissue (Healthy lung tissues)

For this RAG class, include **normal lung histology**:

- Any histology atlas / lecture notes with open access (e.g., university histology sites) describing: alveoli, bronchioles, type I/II pneumocytes, elastic stroma, normal bronchial epithelium. If you have PDFs from teaching material, index them.
- Pair these with short explanatory text summarizing how benign lung parenchyma appears vs malignant patterns (loss of architecture, atypia, invasion); WHO and pathology texts above also provide comparisons.