**生物医药合作项目开发**

**研究方向： 哮喘相关 ;**

**委托人： 赵蓓蓓 ;**

**受托人： 杭州铂赛生物科技有限公司 .**

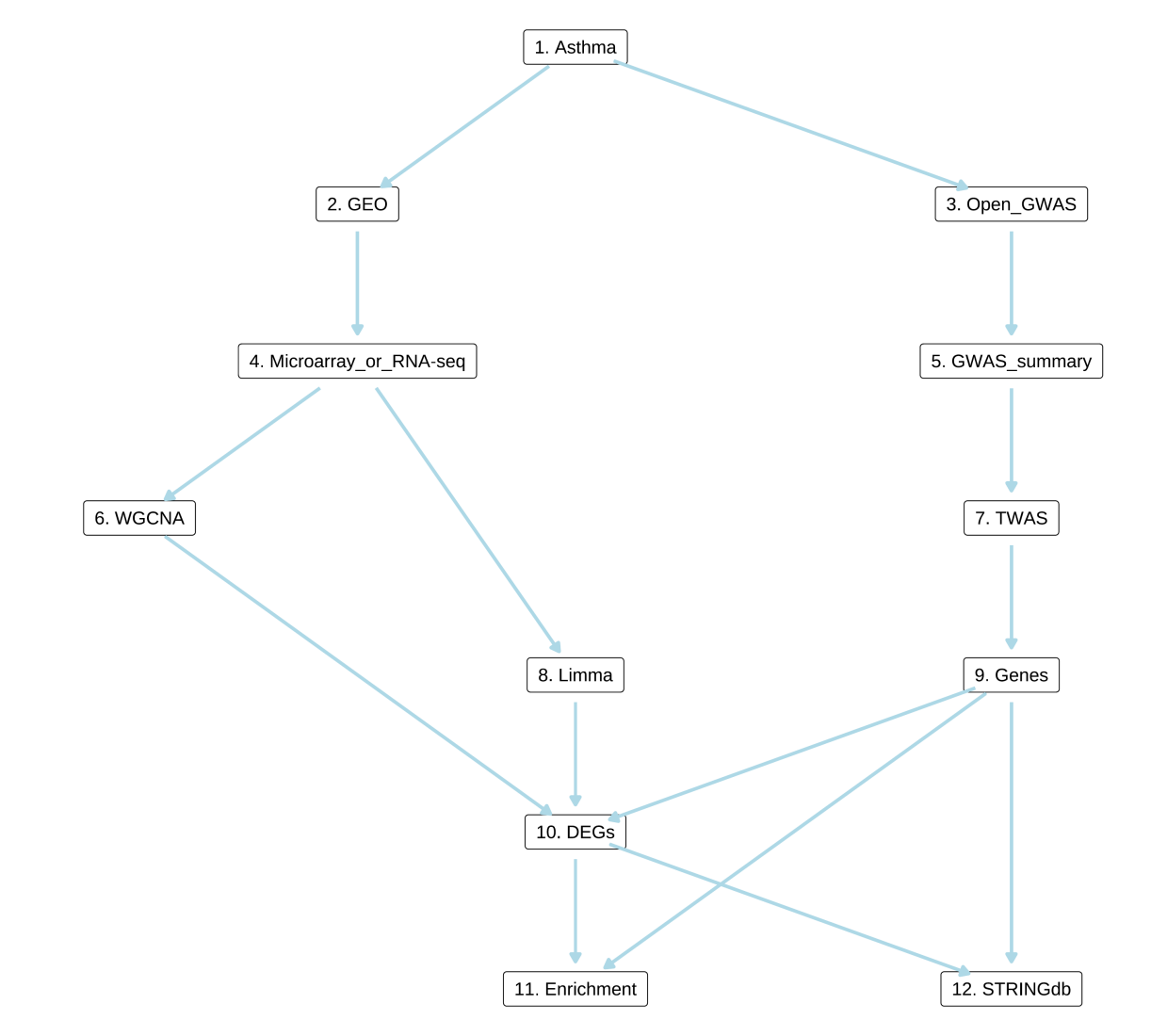
# 1 研究背景

哮喘是影响全球儿童和成人的最常见呼吸道疾病之一，其表型多样，其潜在致病机制尚不明确。随着基因组测序技术的进步，人们做出了科学努力来解释和预测哮喘的复杂性和异质性，全基因组关联研究 (GWAS) 迅速成为首选的研究方法 (2021, **IF:4.9**, Q2, International journal of molecular sciences)1。

TWAS （Transcriptome-Wide Association Study，全转录组关联研究）是一种强大的方法，用于探索基因表达与表型之间的关联，为基因组学研究和遗传病因探索提供了重要工具。它结合了 GWAS 的遗传变异数据和功能基因组数据，能够识别潜在的功能基因，为疾病机制研究和药物开发提供重要线索。

将 TWAS (2016, **IF:31.7**, Q1, Nature Genetics)2 和 WGCNA（Weighted Gene Co-Expression Network Analysis，加权基因共表达网络分析）(2008, **IF:2.9**, Q1, BMC Bioinformatics)3以及差异分析结合起来，能够充分挖掘基因表达调控与表型（例如疾病）之间的复杂关系

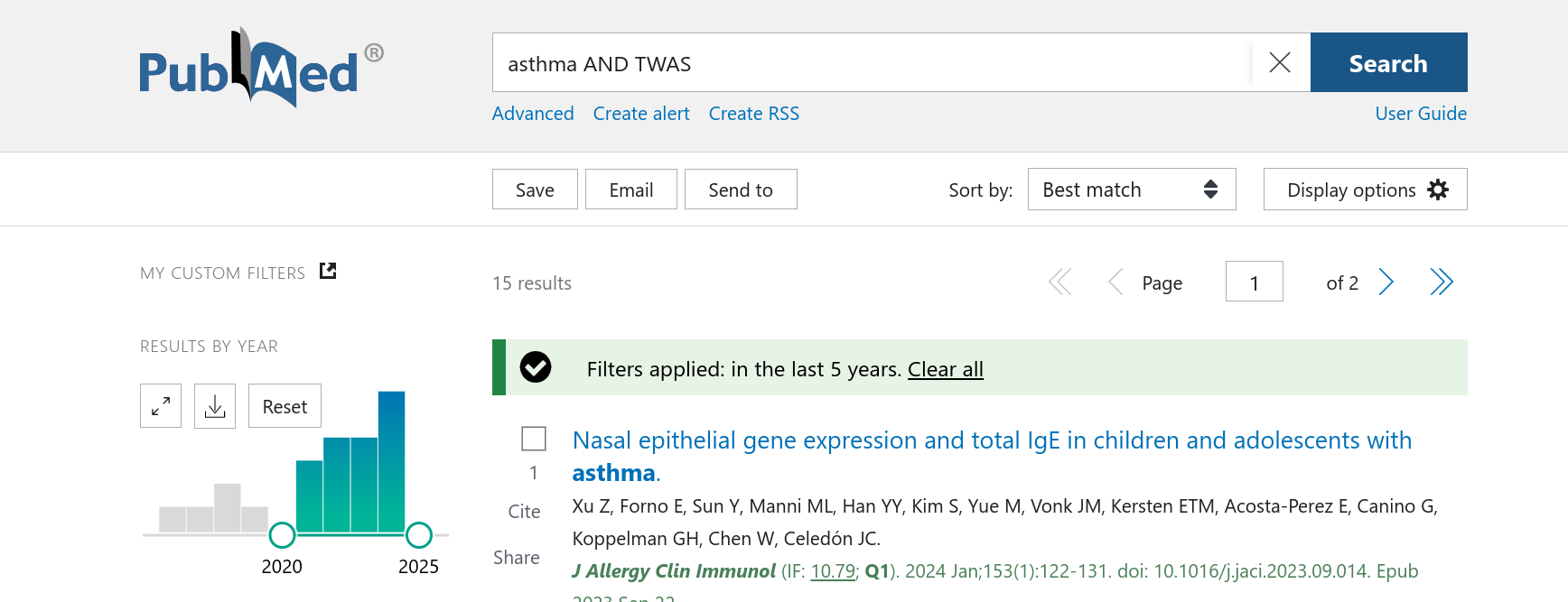
## 1.1 思路



**Fig.** **1** Route

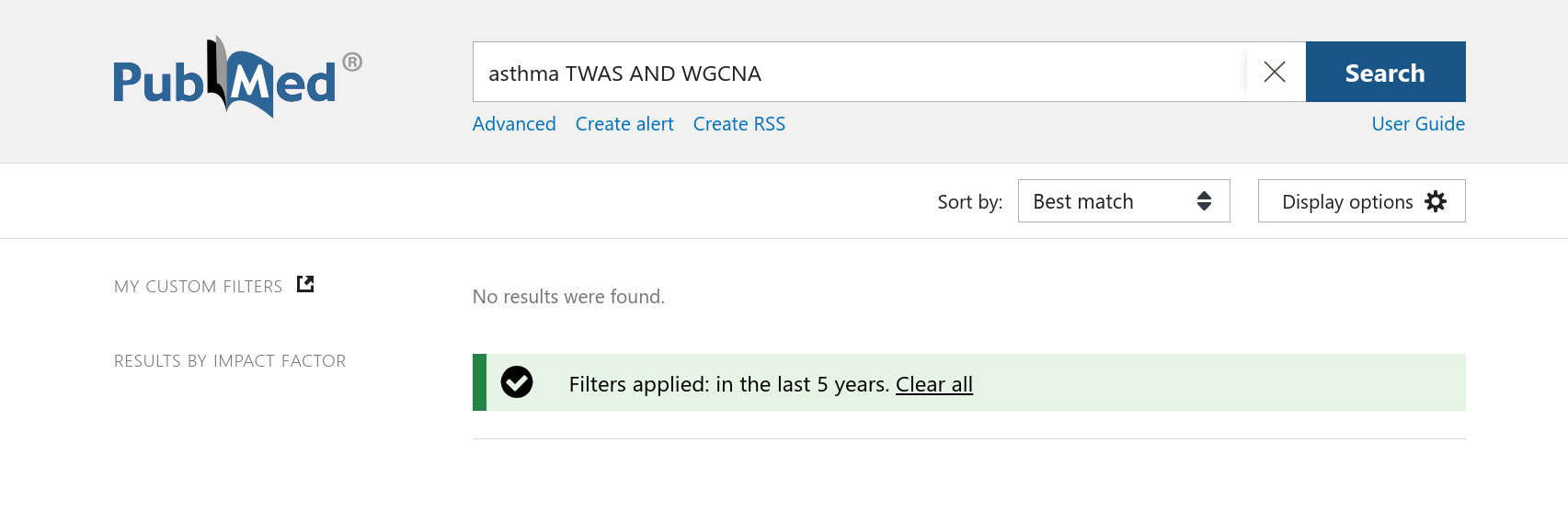
# 2 可行性

## 2.1 以 asthma AND TWAS 搜索文献。



# 3 创新性

## 3.1 以 asthma AND TWAS AND WGCNA 搜索文献。



# 4 GEO 检索方法

## 4.1 数据分析平台

在 Linux pop-os x86\_64 (6.9.3-76060903-generic) 上，使用 R version 4.4.2 (2024-10-31) (<https://www.r-project.org/>) 对数据统计分析与整合分析。

## 4.2 GSE 数据搜索 (Dataset: AS)

使用 Entrez Direct (EDirect) <https://www.ncbi.nlm.nih.gov/books/NBK3837/> 搜索 GEO 数据库 (esearch -db gds)，查询信息为: ((asthma[Description] AND (healthy[Description]) AND ((6:1000[Number of Samples]) AND (GSE[Entry Type]) AND (Homo Sapiens[Organism]))。 以正则匹配，滤除包含 ‘single cell’ 或 ‘scRNA’ 的数据例。仅查询临床样本信息，因此滤除匹配到 ‘cells’, ‘cell type’ 或 ‘cell line’ 的实验数据例。 此外，去除了以特定 Marker 细胞类型为研究对象的数据例 (CD4、CD8 T 细胞等，可能是来源于实验室的数据)。 (注：以上仅为查找合适的 GEO 数据所做的数据筛选，与实际分析无关) 。仅获取类型包含 ‘Expression profiling by high throughput sequencing’ 或 ‘Expression profiling by array’ 的数据例。 以 GEOquery 获取 GSE 数据集 (n=59)。

# 5 参考文献和数据集

## 5.1 GWAS 数据

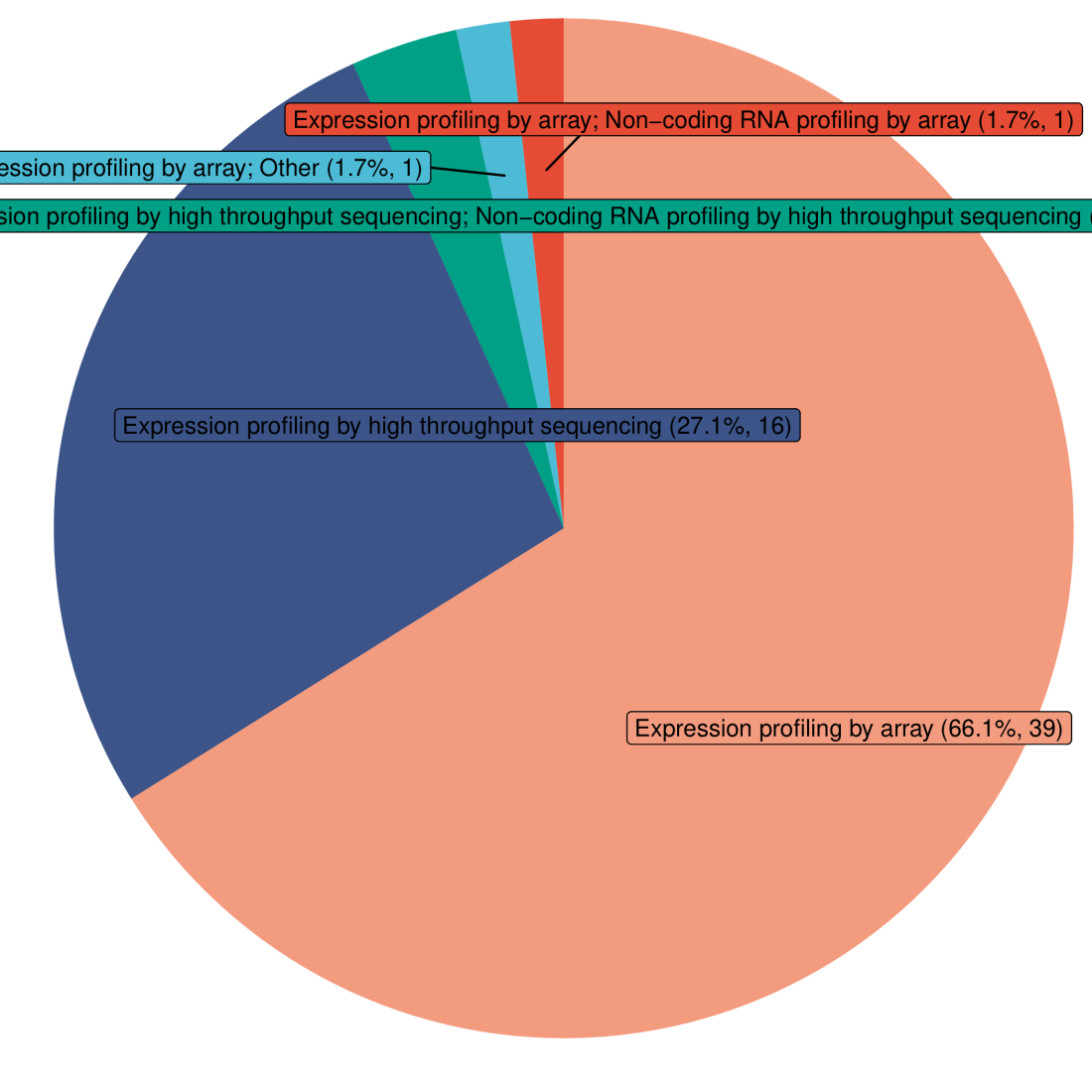
**Tab.** **1** Traits found in Open GWAS

| Id | Trait | Coverage | Ncase | Group ... | Year | Mr | Author | Sex | Qc pri... |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ebi-a-... | Asthma |  | 9209 | Public | 2021 | 1 | Guindo... | NA |  |
| Ebi-a-... | Asthma |  | 9209 | Public | 2021 | 1 | Guindo... | NA |  |
| Ebi-a-... | Asthma |  | 9209 | Public | 2021 | 1 | Guindo... | NA |  |
| Ebi-a-... | Asthma |  | 9209 | Public | 2021 | 1 | Guindo... | NA |  |
| Ebi-a-... | Asthma |  | 56090 | Public | 2021 | 1 | D<U+00... | NA |  |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

## 5.2 GSE 数据搜索 (AS)

以 Entrez Direct (EDirect) 搜索 GEO 数据库 (检索条件见方法章节) 。 可用数据，及其组别为：

* **GSE245608**, **Type**: RNA-seq
  + Control (n = 12)
  + MAA (n = 11)
  + SAA (n = 9)
* **GSE256534**, **Type**: RNA-seq
  + Cont (n = 13)
  + Pat (n = 19)
* **GSE252878**, **Type**: RNA-seq
  + IL-13 (n = 9)
  + No\_stim (n = 18)
* **GSE252877**, **Type**: RNA-seq
  + Asthma (n = 4)
  + Healthy (n = 4)
* **GSE252876**, **Type**: RNA-seq
  + IL-13 (n = 9)
  + No\_stim (n = 18)
* **GSE226825**, **Type**: RNA-seq
  + cells (n = 436)
* **GSE222129**, **Type**: RNA-seq
  + Asthma (n = 27)
  + Healthy (n = 27)
* **GSE207751**, **Type**: RNA-seq
  + H (n = 14)
  + M (n = 19)
  + S (n = 23)
* **GSE184382**, **Type**: Microarray
  + no (n = 23)
  + yes (n = 16)
* **GSE195599**, **Type**: RNA-seq
  + Asthma\_rep (n = 4)
  + Healthy\_rep (n = 3)
* **GSE182798**, **Type**: Microarray
  + NA (n = 102)
  + Nasal\_biopsy (n = 80)
* **GSE182797**, **Type**: Microarray
  + AAD (n = 13)
  + AND (n = 15)
  + APD (n = 17)
  + Ctrl (n = 21)
  + IEI (n = 14)
* **GSE182503**, **Type**: Microarray
  + AAD (n = 16)
  + AND (n = 17)
  + APD (n = 17)
  + Ctrl (n = 21)
  + Ctrl\_II (n = 15)
  + IEI (n = 16)
* **GSE147881**, **Type**: Microarray
  + healthy (n = 13)
  + mild/moderate\_asthma (n = 18)
  + severe\_asthma (n = 42)
* **GSE147880**, **Type**: Microarray
  + healthy\_control (n = 10)
  + mild/moderate\_asthma (n = 18)
  + severe\_asthma (n = 16)
* **GSE147878**, **Type**: Microarray
  + healthy (n = 13)
  + mild/moderate\_asthma (n = 18)
  + severe\_asthma (n = 42)
* **GSE143303**, **Type**: Microarray
  + healthy (n = 13)
  + Severe\_asthma (n = 47)
* **GSE179156**, **Type**: Microarray
  + n/a (n = 29)
  + off (n = 38)
  + on (n = 19)
* **GSE143192**, **Type**: Microarray; Non-coding Microarray
  + normal (n = 2)
  + patients (n = 1)
  + S (n = 5)
* **GSE167225**, **Type**: Microarray
  + allergy (n = 16)
  + allergy+asthma (n = 14)
  + healthy (n = 10)
* **GSE165934**, **Type**: Microarray
  + asthma (n = 10)
  + control (n = 9)
* **GSE148004**, **Type**: Microarray
  + AE (n = 9)
  + CN (n = 7)
  + HC (n = 9)
* **GSE134544**, **Type**: Microarray
  + Healthy\_Control\_Blood\_Donor\_BD (n = 13)
  + Healthy\_Control\_Blood\_Donor\_BD105933\_Replicate (n = 2)
  + Healthy\_Control\_Blood\_Donor\_BD106691\_Replicate (n = 2)
  + Healthy\_Control\_Blood\_Donor\_BD110992\_Replicate (n = 2)
  + Healthy\_Control\_Blood\_Donor\_BD155612\_Replicate (n = 2)
  + Patient\_X (n = 45)
  + Patient\_X10 (n = 3)
  + Patient\_X11 (n = 4)
  + Patient\_X12 (n = 4)
  + Patient\_X13 (n = 4)
  + Patient\_X14 (n = 4)
  + Patient\_X15 (n = 4)
  + Patient\_X16 (n = 4)
  + Patient\_X17 (n = 4)
  + Patient\_X18 (n = 4)
  + Patient\_X19 (n = 4)
  + Patient\_X2 (n = 4)
  + Patient\_X20 (n = 4)
  + Patient\_X21 (n = 3)
  + Patient\_X22 (n = 4)
  + Patient\_X23 (n = 4)
  + Patient\_X25 (n = 4)
  + Patient\_X26 (n = 4)
  + Patient\_X27 (n = 4)
  + Patient\_X28 (n = 4)
  + Patient\_X29 (n = 4)
  + Patient\_X30 (n = 4)
  + Patient\_X31 (n = 4)
  + Patient\_X32 (n = 4)
  + Patient\_X33 (n = 4)
  + Patient\_X34 (n = 4)
  + Patient\_X35 (n = 4)
  + Patient\_X36 (n = 4)
  + Patient\_X37 (n = 4)
  + Patient\_X38 (n = 4)
  + Patient\_X39 (n = 4)
  + Patient\_X40 (n = 4)
  + Patient\_X41 (n = 4)
  + Patient\_X42 (n = 4)
  + Patient\_X43 (n = 4)
  + Patient\_X44 (n = 4)
  + Patient\_X45 (n = 4)
  + Patient\_X46 (n = 4)
  + Patient\_X47 (n = 4)
  + Patient\_X48 (n = 4)
  + Patient\_X49 (n = 4)
  + Patient\_X5 (n = 2)
  + Patient\_X50 (n = 4)
  + Patient\_X6 (n = 3)
  + Patient\_X7 (n = 3)
  + Patient\_X9 (n = 3)
* **GSE152004**, **Type**: RNA-seq
  + HR (n = 686)
  + SJ (n = 9)
* **GSE152729**, **Type**: RNA-seq; Non-coding RNA-seq
  + Healthy (n = 8)
  + severe\_asthmatic (n = 12)
* **GSE151631**, **Type**: RNA-seq
  + healthy\_control (n = 7)
  + Keratoconus (n = 19)
* **GSE101720**, **Type**: RNA-seq
  + B.AR (n = 7)
  + B.C (n = 9)
  + B.R (n = 10)
  + N.AR (n = 7)
  + N.C (n = 9)
  + N.R (n = 10)
* **GSE132006**, **Type**: Microarray
  + Allergic\_Asthmatic (n = 10)
  + Healthy\_Control (n = 10)
* **GSE110551**, **Type**: Microarray
  + W (n = 117)
  + Z (n = 39)
* **GSE137268**, **Type**: Microarray
  + Sputum\_Asthma (n = 54)
  + Sputum\_Healthy (n = 15)
* **GSE128027**, **Type**: RNA-seq
  + 09 (n = 1)
  + 10\_007\_IL (n = 1)
  + 11\_007\_IL (n = 1)
  + 12 (n = 1)
  + 13 (n = 1)
  + 14\_006\_IL (n = 1)
  + 15\_006\_IL (n = 1)
  + 16 (n = 1)
  + 17 (n = 1)
  + 18\_002\_IL (n = 1)
  + 19\_002\_IL (n = 1)
  + 20 (n = 1)
  + 21 (n = 1)
  + 22\_018\_IL (n = 1)
  + 23\_018\_IL (n = 1)
  + 24 (n = 1)
  + 25 (n = 1)
  + 26\_019\_IL (n = 1)
  + 27\_019\_IL (n = 1)
  + 28 (n = 1)
  + GMCSF (n = 2)
  + IL3 (n = 2)
  + IL5 (n = 2)
  + neg\_control (n = 2)
* **GSE127773**, **Type**: RNA-seq
  + Asthmatic (n = 40)
  + Healthy (n = 24)
* **GSE117038**, **Type**: RNA-seq; Non-coding RNA-seq
  + blood\_\_health\_\_lncRNA\_mRNA\_rep (n = 3)
  + blood\_Eos\_\_lncRNA\_mRNA\_rep (n = 3)
  + blood\_Neu\_\_lncRNA\_mRNA\_rep (n = 6)
* **GSE120855**, **Type**: Microarray
  + Mononuclear\_Leukocytes (n = 10)
* **GSE114669**, **Type**: Microarray
  + Healthy\_controls (n = 12)
  + Severe\_asthma (n = 10)
* **GSE85307**, **Type**: Microarray; Other
  + Subject (n = 155)
  + Subject\_1050601-M,\_microRNA,\_Normal\_Pregnancy,\_Replicate (n = 2)
  + Subject\_3017557-M,\_microRNA,\_Preeclampsia\_Pregnancy,\_Replicate (n = 2)
* **GSE97668**, **Type**: RNA-seq
  + Asthmatic (n = 15)
  + Non-asthmatic (n = 18)
* **GSE17905**, **Type**: Microarray
  + large\_airways,\_non-smoker (n = 21)
  + large\_airways,\_smoker (n = 31)
  + small\_airways,\_non-smoker (n = 41)
  + Small\_airways,\_non-smoker (n = 6)
  + small\_airways,\_smoker (n = 56)
  + small\_airways,\_smoker\_058,\_MAS (n = 1)
  + small\_airways,\_smoker\_063,\_MAS (n = 1)
* **GSE64913**, **Type**: Microarray
  + Healthy\_Central\_Patient (n = 23)
  + Healthy\_Peripheral\_Patient (n = 19)
  + Severe\_Central\_Patient (n = 13)
  + Severe\_Peripheral\_Patient (n = 15)
* **GSE69683**, **Type**: Microarray
  + Healthy,\_non-smoking (n = 87)
  + Moderate\_asthma,\_non-smoking (n = 77)
  + Severe\_asthma,\_non-smoking (n = 246)
  + Severe\_asthma,\_smoking (n = 88)
* **GSE83233**, **Type**: (scRNA-seq) Microarray
  + Budesonide (n = 11)
  + Placebo (n = 11)
* **GSE65239**, **Type**: Microarray
  + Dermatologic\_Disease (n = 3)
  + Healthy\_Control (n = 3)
  + Parasitic\_Disease (n = 3)
  + Pulmonary\_Aspergillosis (n = 4)
  + Severe\_Asthma (n = 4)
* **GSE65204**, **Type**: Microarray
  + 15-02 (n = 22)
  + 15-04 (n = 5)
  + 15-05 (n = 10)
  + 15-06 (n = 5)
  + 15-07 (n = 9)
  + 15-08 (n = 18)
* **GSE41863**, **Type**: Microarray
  + Asthma (n = 47)
  + Control (n = 9)
* **GSE41862**, **Type**: Microarray
  + Asthma (n = 95)
  + Control (n = 21)
* **GSE41861**, **Type**: Microarray
  + Asthma (n = 91)
  + Control (n = 47)
* **GSE74075**, **Type**: Microarray
  + asthmatic (n = 10)
  + healthy (n = 6)
* **GSE61225**, **Type**: Microarray
  + post (n = 29)
  + post\_16 (n = 2)
  + post\_32 (n = 2)
  + post\_48 (n = 2)
  + post\_49 (n = 2)
  + pre (n = 29)
  + pre\_16 (n = 2)
  + pre\_32 (n = 2)
  + pre\_48 (n = 2)
  + pre\_49 (n = 2)
* **GSE67472**, **Type**: Microarray
  + asthma (n = 62)
  + healthy (n = 43)
* **GSE38003**, **Type**: RNA-seq
  + Atopic\_asthma\_airway\_smooth\_muscle\_A (n = 12)
  + Atopic\_healthy\_control\_airway\_smooth\_muscle\_H (n = 6)
  + Non-atopic\_healthy\_control\_airway\_smooth\_muscle\_H (n = 6)
* **GSE54605**, **Type**: Microarray
  + asthma\_patient (n = 6)
  + healthy\_adult (n = 4)
* **GSE46171**, **Type**: Microarray
  + NATURI (n = 14)
  + NATURI\_06\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_07\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_08\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_09\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_11\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_12\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_13\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_14\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_15\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_17\_Nasal\_mucosa\_Healthy\_day (n = 1)
  + NATURI\_21\_Nasal\_mucosa\_Allrg\_Rhin\_day (n = 1)
  + NATURI\_23\_Nasal\_mucosa\_Allrg\_Rhin\_day (n = 1)
  + NATURI\_25\_Nasal\_mucosa\_Allrg\_Rhin\_day (n = 1)
  + NATURI\_27\_Nasal\_mucosa\_Allrg\_Rhin\_day (n = 1)
  + NATURI\_49\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_50\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_51\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_53\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_54\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_59\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_60\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_64\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_65\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_67\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_68\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_73\_Nasal\_mucosa\_AsmNoEx\_day (n = 1)
  + NATURI\_84\_Nasal\_mucosa\_AsmEx\_day (n = 1)
  + NATURI\_87\_Nasal\_mucosa\_AsmEx\_day (n = 1)
  + NATURI\_90\_Nasal\_mucosa\_AsmEx\_day (n = 1)
  + NATURI\_91\_Nasal\_mucosa\_AsmEx\_day (n = 1)
* **GSE41649**, **Type**: Microarray
  + allergic\_asthmatic\_subjects (n = 4)
  + non\_allergic\_non\_asthmatic\_subject,\_without\_any\_stimulation (n = 4)
* **GSE19190**, **Type**: Microarray
  + Healthy (n = 11)
  + NA (n = 21)
  + Rhinitis (n = 14)
  + Rhinitis\_with\_controlled\_asthma (n = 7)
  + Uncontrolled\_asthma (n = 6)
* **GSE23611**, **Type**: Microarray
  + Asthma (n = 49)
  + Healthy (n = 13)
* **GSE20881**, **Type**: Microarray
  + ascending colon biopsy from crohns disease subject PRB (n = 20)
  + ascending colon biopsy from healthy subject GSM (n = 17)
  + descending colon biopsy from crohns disease subject PRB (n = 29)
  + descending colon biopsy from healthy subject GSM (n = 23)
  + sigmoid colon biopsy from crohns disease subject PRB (n = 33)
  + sigmoid colon biopsy from healthy subject GSM (n = 27)
  + terminal ileum biopsy from crohns disease subject PRB (n = 17)
  + terminal ileum biopsy from healthy subject GSM (n = 6)
* **GSE22528**, **Type**: Microarray
  + asthma (n = 5)
  + normal (n = 5)
* **GSE15823**, **Type**: Microarray
  + BL (n = 11)
  + CLE (n = 1)
* **GSE11223**, **Type**: Microarray
  + Normal (n = 73)
  + UC (n = 129)



**Fig.** **2** AS All Gds Type

# Reference

1. Ntontsi, P., Photiades, A., Zervas, E., Xanthou, G. & Samitas, K. Genetics and epigenetics in asthma. *International journal of molecular sciences* **22**, (2021).

2. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245–252 (2016).

3. Langfelder, P. & Horvath, S. WGCNA: An r package for weighted correlation network analysis. *BMC Bioinformatics* **9**, (2008).