**✅ 一、散点图解读：LFC vs Dependency Score**

**坐标轴说明：**

| **轴** | **变量** | **含义** |
| --- | --- | --- |
| x 轴 | crispr\_effect\_LFC | G12C+ vs 非G12C 之间 CRISPR KO 分数差异 |
| y 轴 | crispr\_dependency\_score | KRAS G12C 阳性细胞中对该基因的依赖性概率 |
| 点大小 | crispr\_mean\_effect | KO 后生长抑制强度（负值表示致死） |
| 颜色深浅 | crispr\_consistency | KO 分数在阳性细胞中是否稳定（1最稳定） |

**解读重点：**

1. ✅ **集中分布在 LFC ≈ 0，Dependency ≈ 0~1 区间**：
   * 多数基因在 G12C+ 与非G12C 之间 **KO 分数无显著差异**
   * 所以它们**可能是通用必需基因**，但与 G12C 无直接关联
2. 🔍 **关注左上象限（LFC < 0, Dependency 高）**：
   * 这类基因 KO 后在 G12C+ 细胞系中抑制更强（说明这些基因是 G12C 特异依赖因子）
   * ✅ **候选靶点基因应集中在这个区域**
3. ⚠️ **右侧区域（LFC > 0）**：
   * 表示 KO 后反而“更容易生长”，可能是**抑癌基因**或**负向调控因子**

**✅ 二、热图解读：Top 20 KO LFC 基因的三维特征**

**热图说明：**

* 横轴是三个 CRISPR 特征维度：
  + KO 得分均值 crispr\_mean\_effect
  + 一致性 crispr\_consistency
  + 依赖概率 crispr\_dependency\_score
* 纵轴是 LFC 最大的 Top 20 基因

**解读示例：**

| **基因** | **KO效果** | **一致性** | **依赖性** | **解读** |
| --- | --- | --- | --- | --- |
| **ZNHIT2** | -1.68 | 0.65 | **0.99** | G12C+ 中 KO 致死性高，依赖度极高 ✅ 极佳靶点 |
| **SNAPC2** | -1.47 | 0.58 | 0.96 | KO 效果强，依赖性高，波动略大，仍可作为备选靶点 |
| **EP300** | 0.01 | 0.77 | 0.09 | 几乎没有 KO 效果，不依赖 ❌ 排除 |

**一、KO 效果（crispr\_mean\_effect）**

**💡 全称：平均 KO 后的生长效应评分**

这个分数来源于 **DepMap 的 CRISPR KO 实验**：  
他们用 CRISPR 把某个基因“敲除”（knock out），然后观察细胞是否还能活下去。

| **分数值** | **含义** | **示例** |
| --- | --- | --- |
| **接近 0** | 该基因不是必需的 | 敲了没啥反应，细胞正常活着 |
| **负值（<0）** | KO 后细胞长不动了，说明**这个基因是必需的** | TP53, RPA3 等 |
| **越小越重要** | KO 后死亡 → 潜在靶点 | -0.4, -1.2 等 |

👉KO某个基因后细胞死亡（因药）说明这个基因可能是维持耐药性所必需的，一旦 KO，细胞就没法维持耐药状态，甚至无法生存。”

**✅ 二、一致性（crispr\_consistency = 1 - std）**

**💡 全称：该基因在所有 G12C 阳性细胞中的 KO 分数的一致性**

这是用来衡量：

**某个基因的 KO 分数在所有 KRAS G12C 细胞系中波动大不大？**

| **指标** | **含义** |
| --- | --- |
| **std 越小，一致性越高** | KO 后反应在每个细胞系中都差不多 |
| **1 - std 越接近 1 → 越一致** | 稳定性强，更值得信赖 |

**举例：**

| **基因** | **KO 平均分** | **一致性** | **解读** |
| --- | --- | --- | --- |
| ZNHIT2 | -1.68 | 0.96 | KO 后几乎每个 G12C 细胞都死亡，一致性强 ✅ |
| CDK7 | -0.41 | 0.50 | 有的细胞死，有的不死，可能是背景差异大 ⚠️ |

👉 这能帮助我们过滤掉那些“表现不稳定”的候选靶点。

**✅ 三、依赖性（crispr\_dependency\_score）**

**💡 来源：DepMap 用 Bayesian 模型推算出的“细胞是否依赖某个基因”的概率**

它的原始含义是：

如果一个细胞在 CRISPR KO 某基因后活得很差，那它就是**依赖这个基因**。

| **分数值** | **含义** | **示例** |
| --- | --- | --- |
| 接近 **1.0** | 依赖性极强 → 这个基因是“成瘾基因” ✅ |  |
| 接近 **0.0** | 没有依赖性，不 KO 也无所谓 |  |

这个指标会结合 KO 效果 + 统计模型判断是否属于“common essential”。

👉 在你项目中，**我们只关注 G12C 阳性细胞的依赖评分均值**，以识别 G12C 特异依赖因子。

**🔁 三者关系总结为：**

| **维度** | **意义** | **推荐系统作用** |
| --- | --- | --- |
| KO 效果 | 敲掉是否导致细胞死亡 | 用于衡量直接功能重要性 |
| 一致性 | 敲掉这个基因在不同细胞中反应是否一致 | 衡量数据稳定性（可信度） |
| 依赖性 | 细胞是否“成瘾”于这个基因 | 衡量靶点潜力 |

**✅ 三、结合推荐系统的策略**

基于这两图

| **策略** | **实施方式** |
| --- | --- |
| 🎯 构建 Top 候选因子列表 | 基于 LFC < -1, consistency > 0.6, dependency > 0.8 |
| 🧬 富集分析 Top 20 基因 | 使用热图基因做 GO / KEGG 分析，追溯通路机制 |
| ⚙️ 输入推荐系统字段 | 将这些基因加入 app\_data\_kras.csv 并提升权重（高评分推荐） |
| 📈 构建机制模型 | 追踪这些基因的上下游通路（如 ZNHIT2 在转录复合物中角色） |

附：

| **特征名** | **类型** | **推荐系统价值** | **解释作用** |
| --- | --- | --- | --- |

|  |  |  |  |
| --- | --- | --- | --- |
| crispr\_dependency\_score | 连续数值（0~1） | ✅ 衡量“成瘾性” | 基因越被依赖 → 越可能是驱动因素 |

|  |  |  |  |
| --- | --- | --- | --- |
| crispr\_mean\_effect | 连续值（多为负） | ✅ KO 后生长抑制强 → 潜在弱点 | 越负越说明 KO 致死或抑制生长 |

|  |  |  |  |
| --- | --- | --- | --- |
| crispr\_consistency | 连续值（0~1） | ✅ 越接近 1 越稳定可信 | 可用于置信度加权或筛选 |

|  |  |  |  |
| --- | --- | --- | --- |
| crispr\_effect\_LFC | 连续值（差值） | ✅ G12C+ vs 非G12C 的差异性 | 判断特异性效应 |

|  |  |  |  |
| --- | --- | --- | --- |
| crispr\_effect\_pval | 连续值（0~1） | ⚠️ 可选（小为好） | 显著性检验，越小越显著 |