对单倍体基因组，设：

CDS比例: a ∈ (0,1)

CDS区影响权重pre bp: m

平均移码影响值（pre CDS）: s

标准化调控区比例（含启动子、增强子加权值）： b ∈ (0,1)

调控区影响权重pre bp: n

基因组大小： g

sv不确定度： v ∈ (0,1)

复合sv拆分系数： c ∈ (0,1)

平均CDS长度 d ∈ (0,1)

则有：

a+b<1

SNP：

∀ 位点xi，其在CDS区的概率为a，在调控区的概率为b，

故产生的影响为 a∙m+b∙n

对x个SNP，∑，有x(am+bn);

按基因组大小归一化，得：x(am+bn)/g.

Indel:

∀ 长度l，其在CDS区的平均长度为al，在调控区的平均长度为bl，

∵每个位于CDS区内的片段有2/3的概率造成移码（al<d）

∴移码次数为2al/3d.

而CDS区有ga/d个，

故移码影响为 (2al/3d \* s)/(ga/d)=2ls/3g .

故，总影响：2ls/3g + al/3g + bln/g。

对 总长为 l 的indel，∑，

得：l(2s + a + 3bn)/3g。

SV (由于有复合的，只能每个单独算):

sv不确定度是指文件中报道长为l的区域为包含某类sv的区域，但实际上其中生物学sv的长度只有vl < l。

对复合类型的sv，分别计算，再乘以c加权相加。根据复合数量N设置c，使1 < Nc < N就可以了。

∀一个SV，覆盖CDS区部分长为vla，包括 vla/d 个CDS。

转座：

转座元件长lv/2

倒位区、重复区、Insertion/Deletion区长 lv

将长度代入SV的公式，得：

转座： lv(2s + a + 3bn)/6g。

倒位、重复、Insertion/Deletion： lv(2s + a + 3bn)/3g。

复合的SV，累加后乘以c。