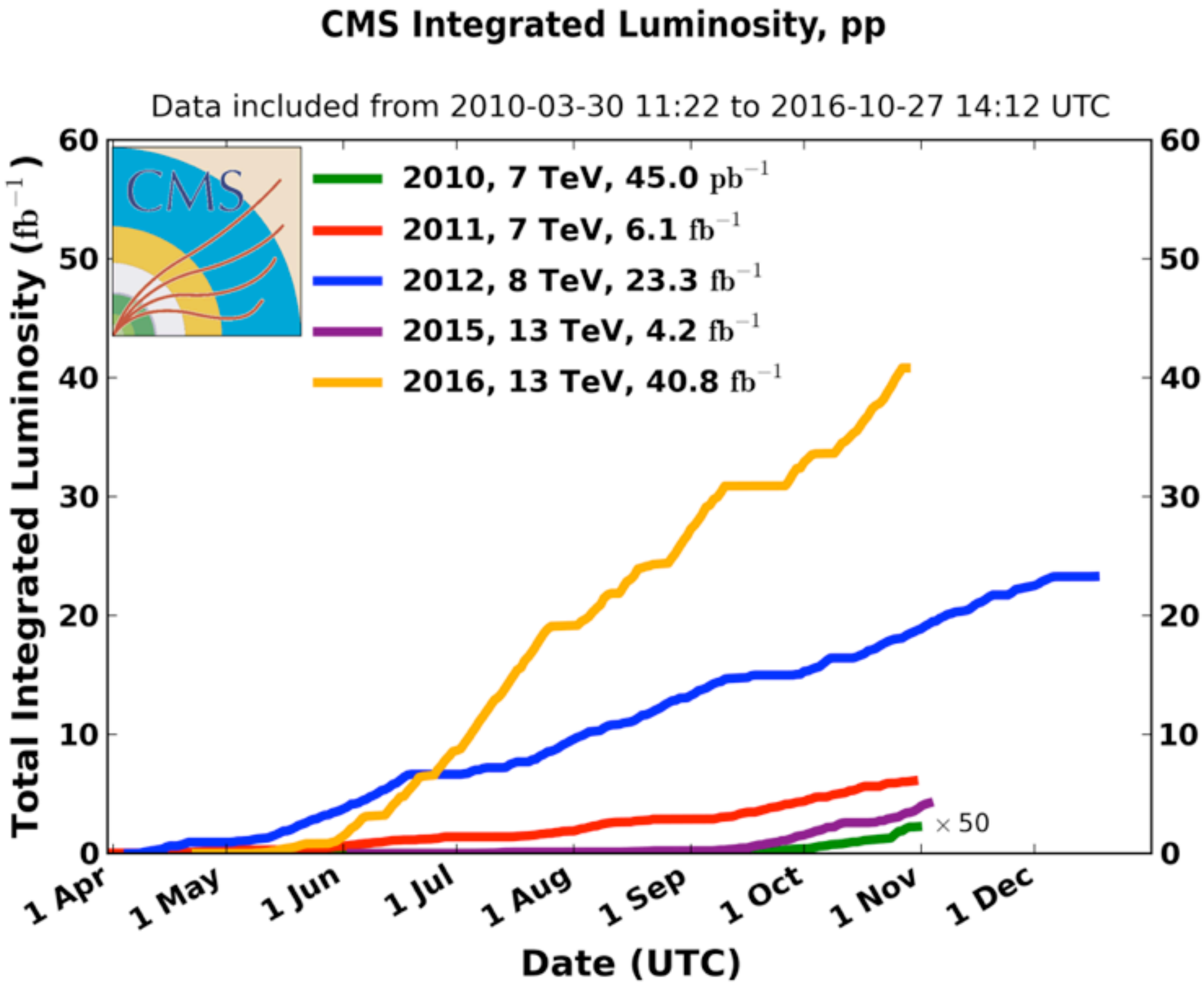
Are there hints for new physics in the Moriond data?

David Shih
NHETC, Rutgers University

CERN-CKC Jeju Workshop
June 1, 2017

Buckley, Feld, Macaluso, Monteux & DS 1610.08059
Buckley, Monteux & DS 1611.05873
Asadi, Buckley, DiFranzo, Monteux & DS in progress

Since 2010, the LHC has been performing spectacularly.

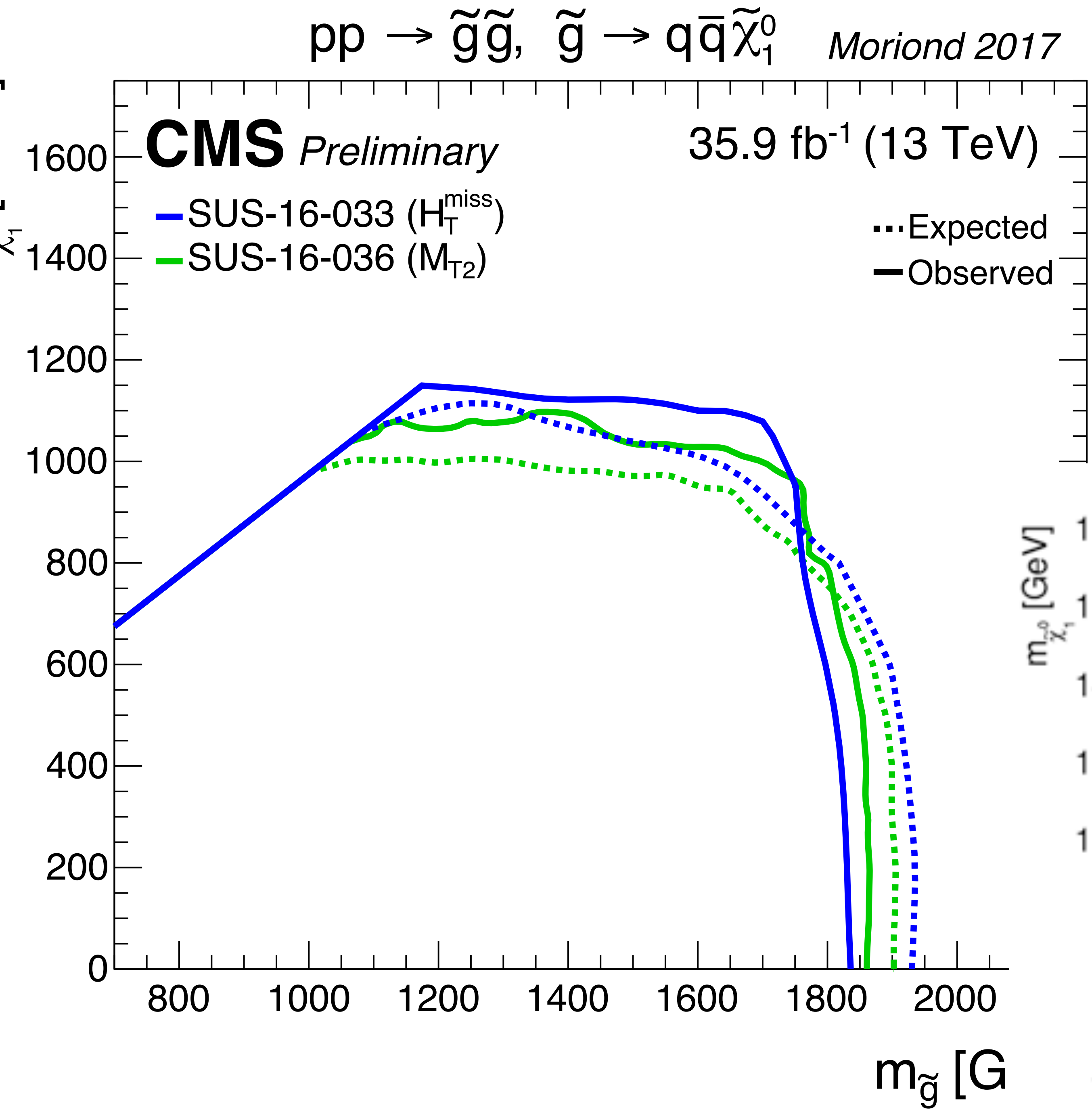


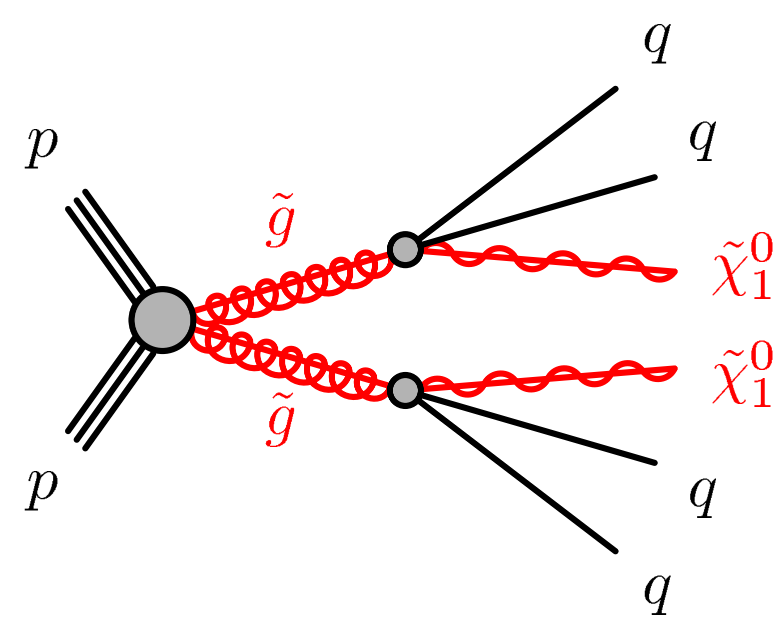
Recently, we reached important milestones:

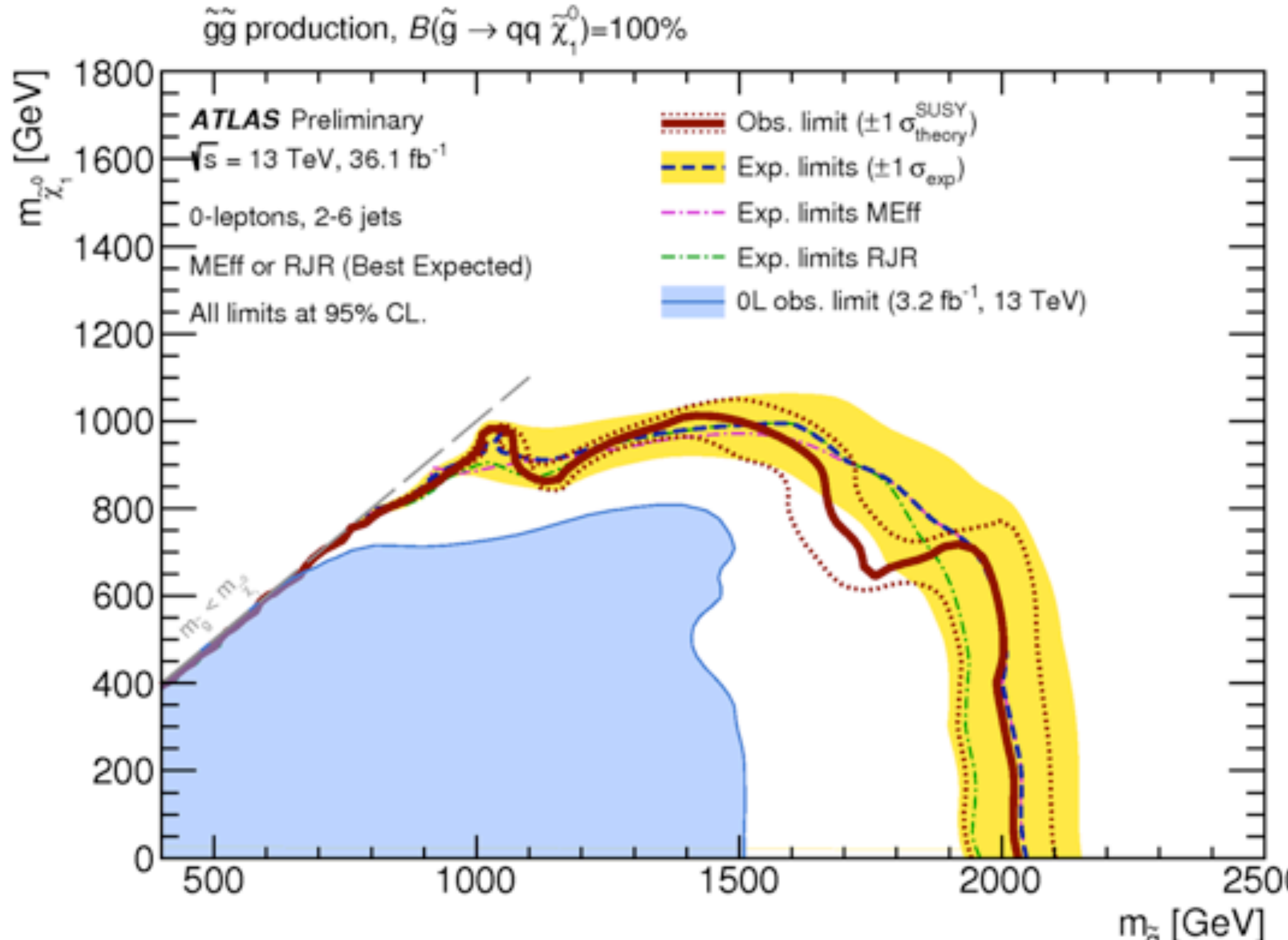
~10/fb at 13 TeV

~40/fb at 13 TeV

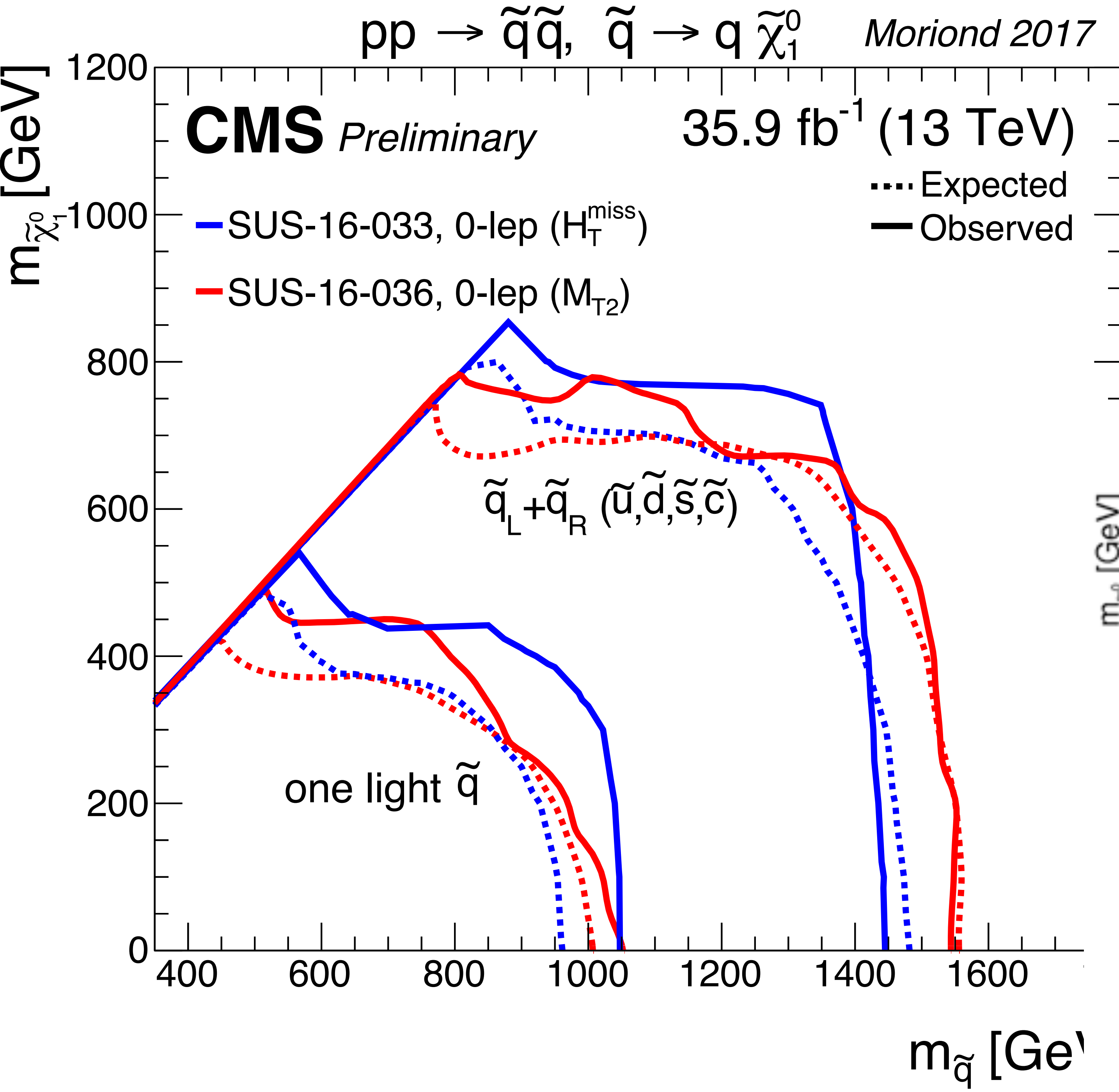
Where is(n't) the new physics??

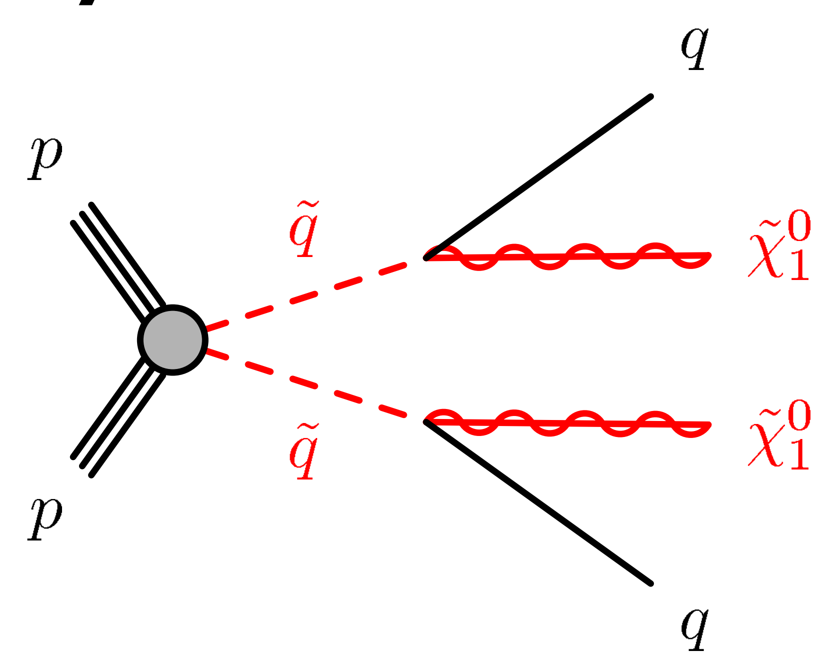


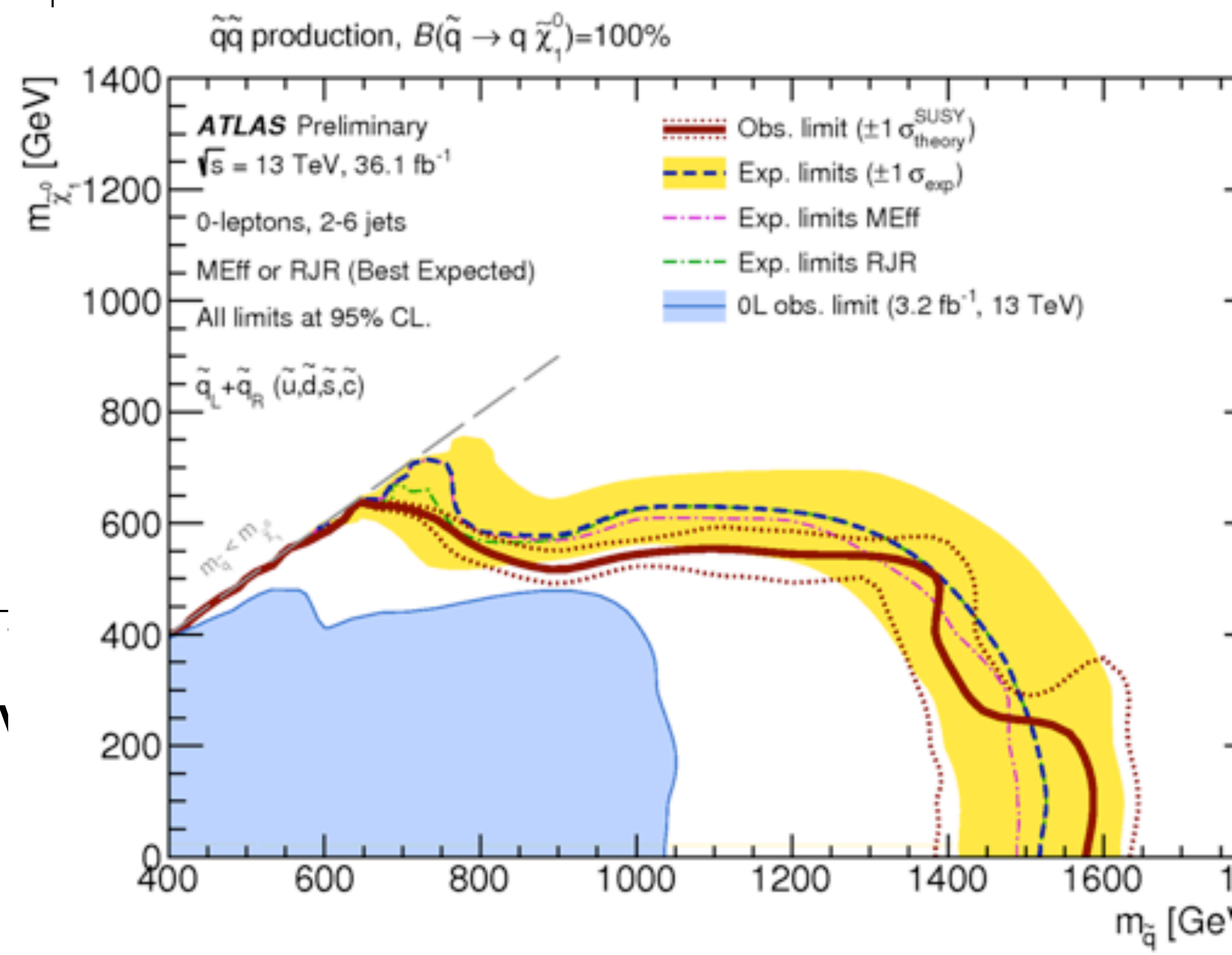




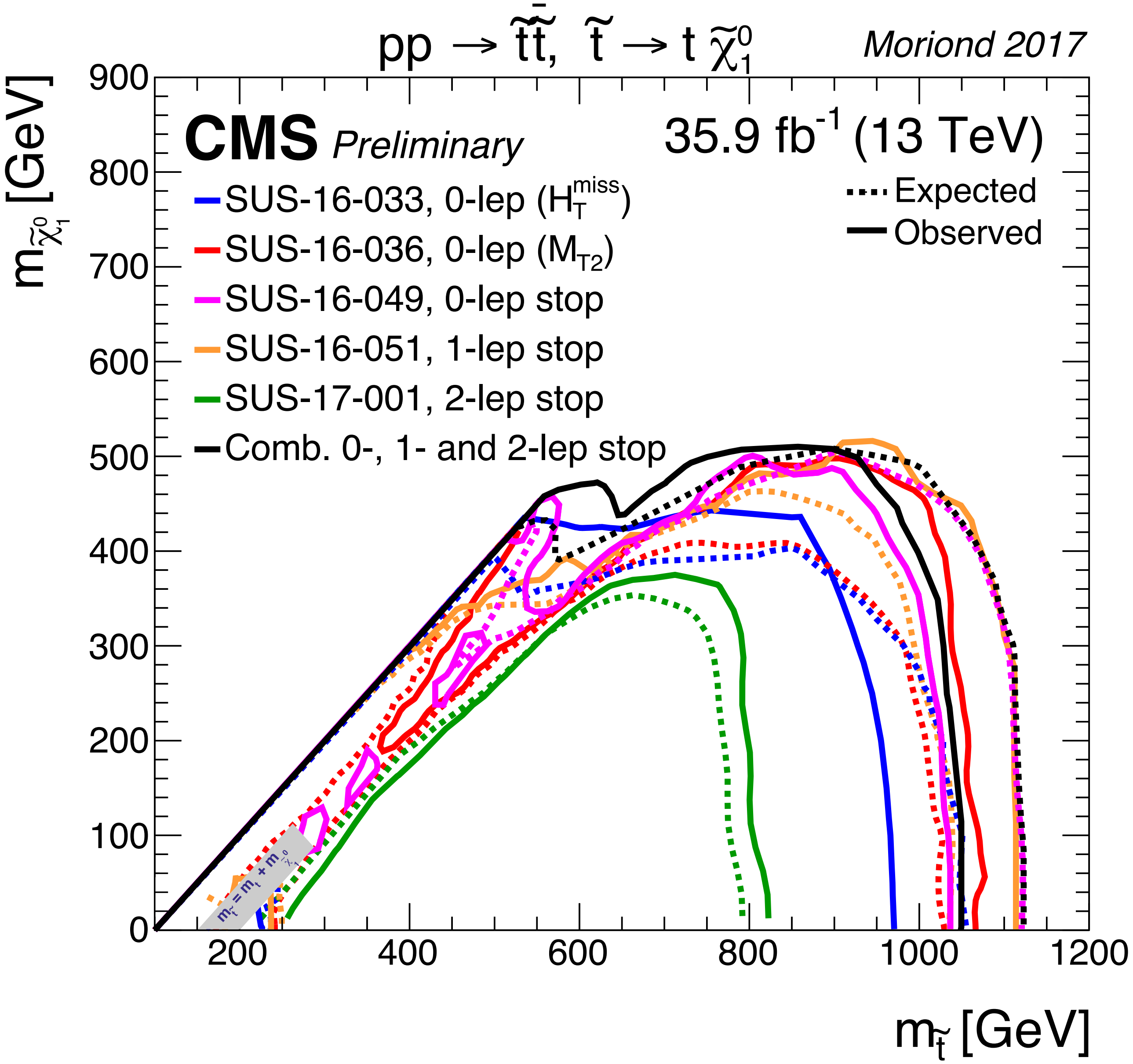
Where is(n't) the new physics??

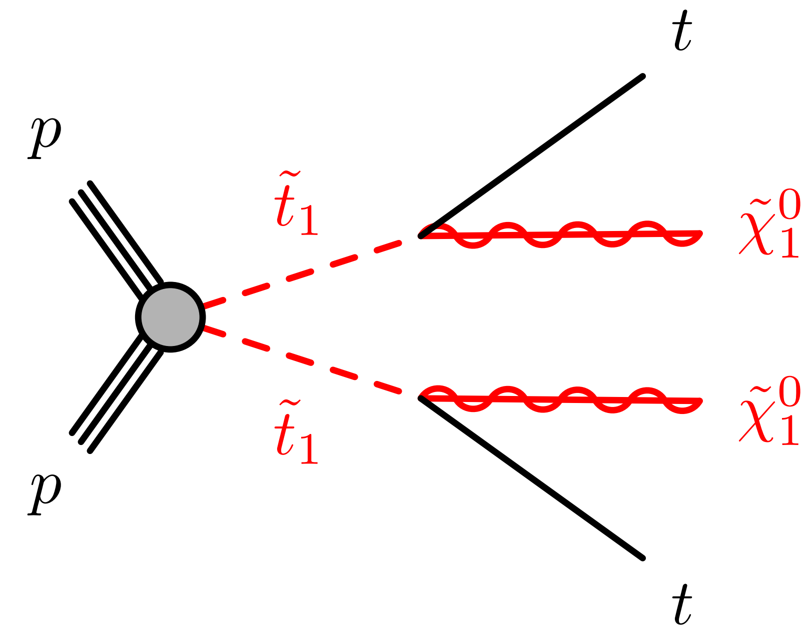


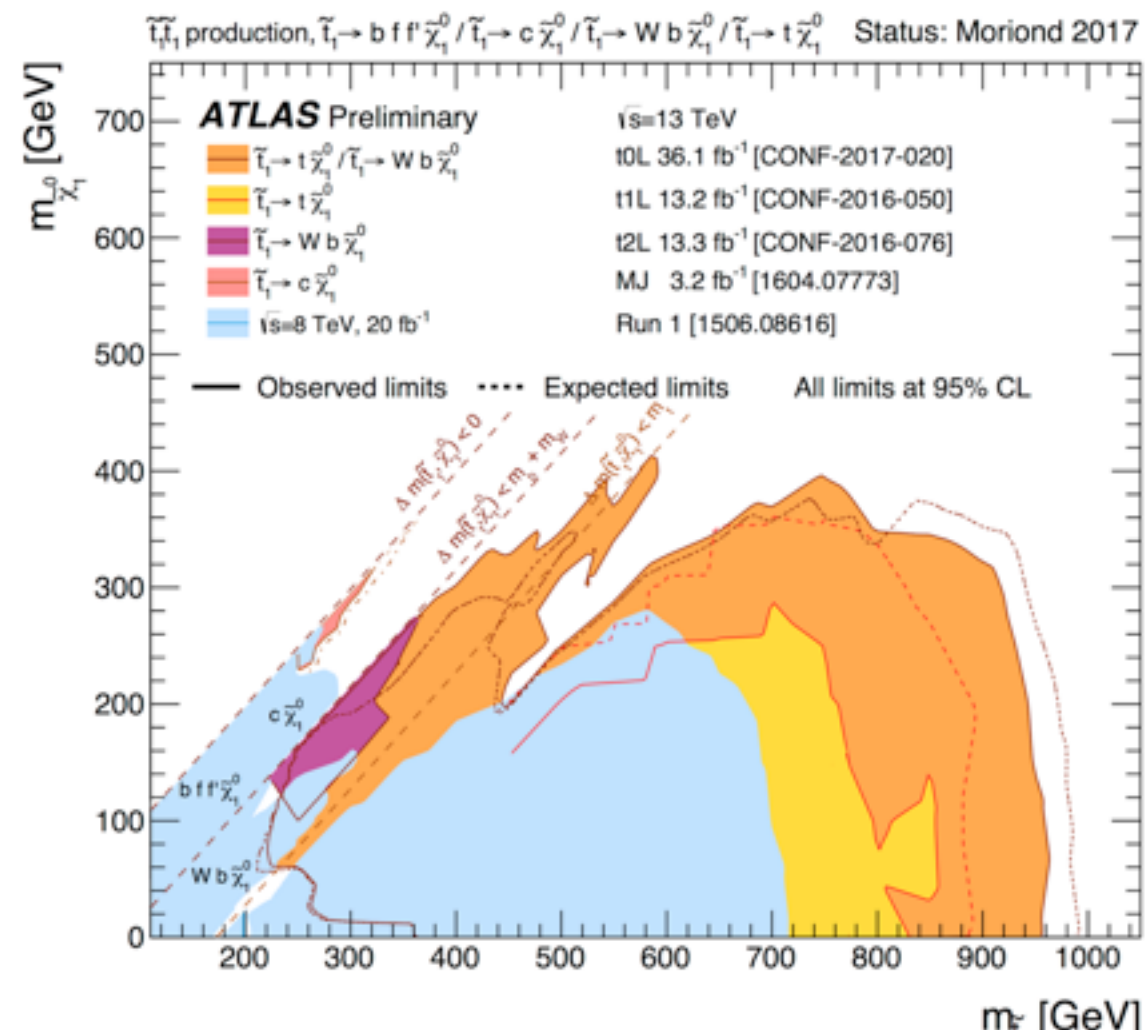




Where is(n't) the new physics??

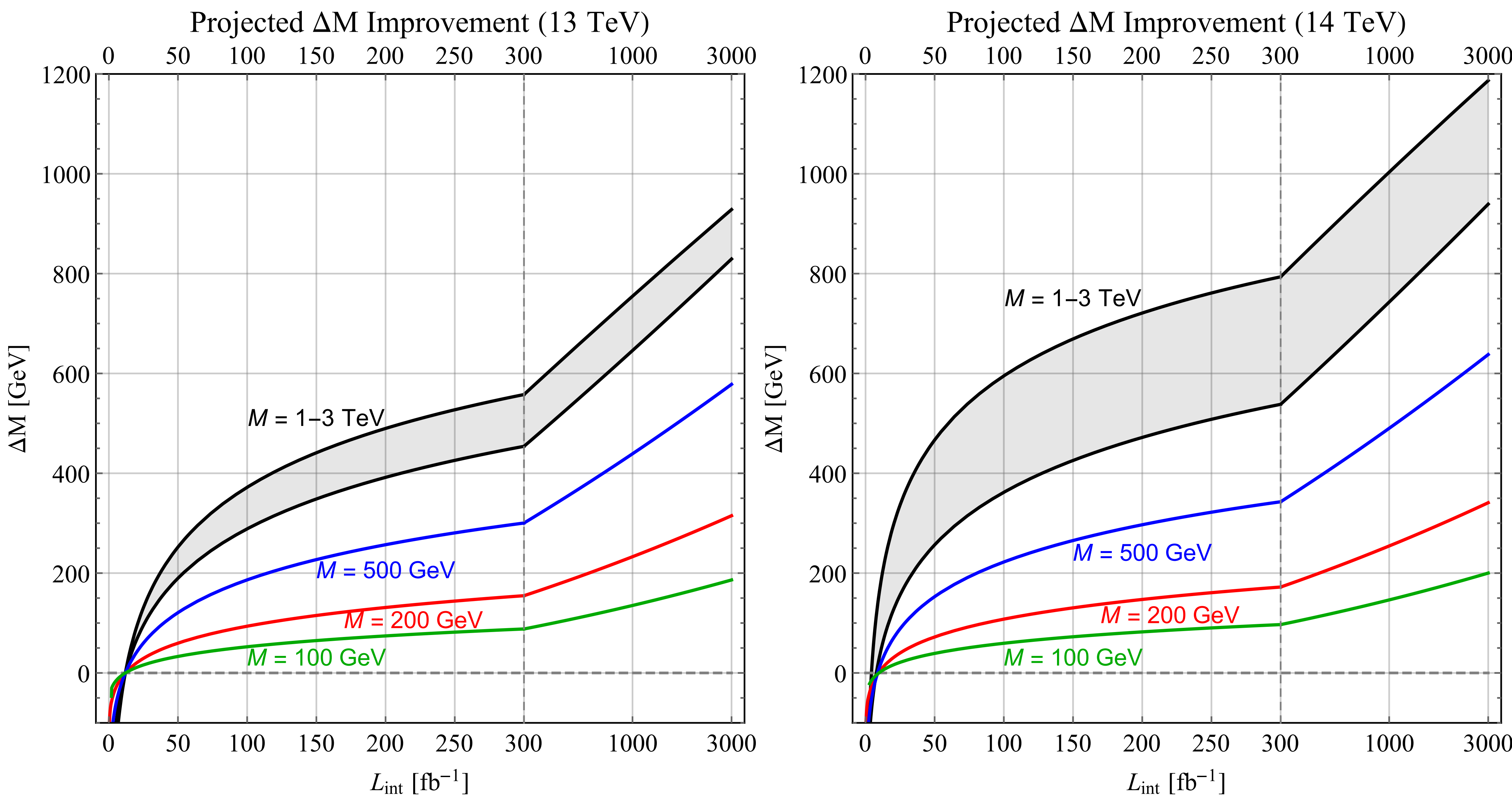






Growth in limits from ICHEP to Moriond (~100-200 GeV) were in line with expectations.

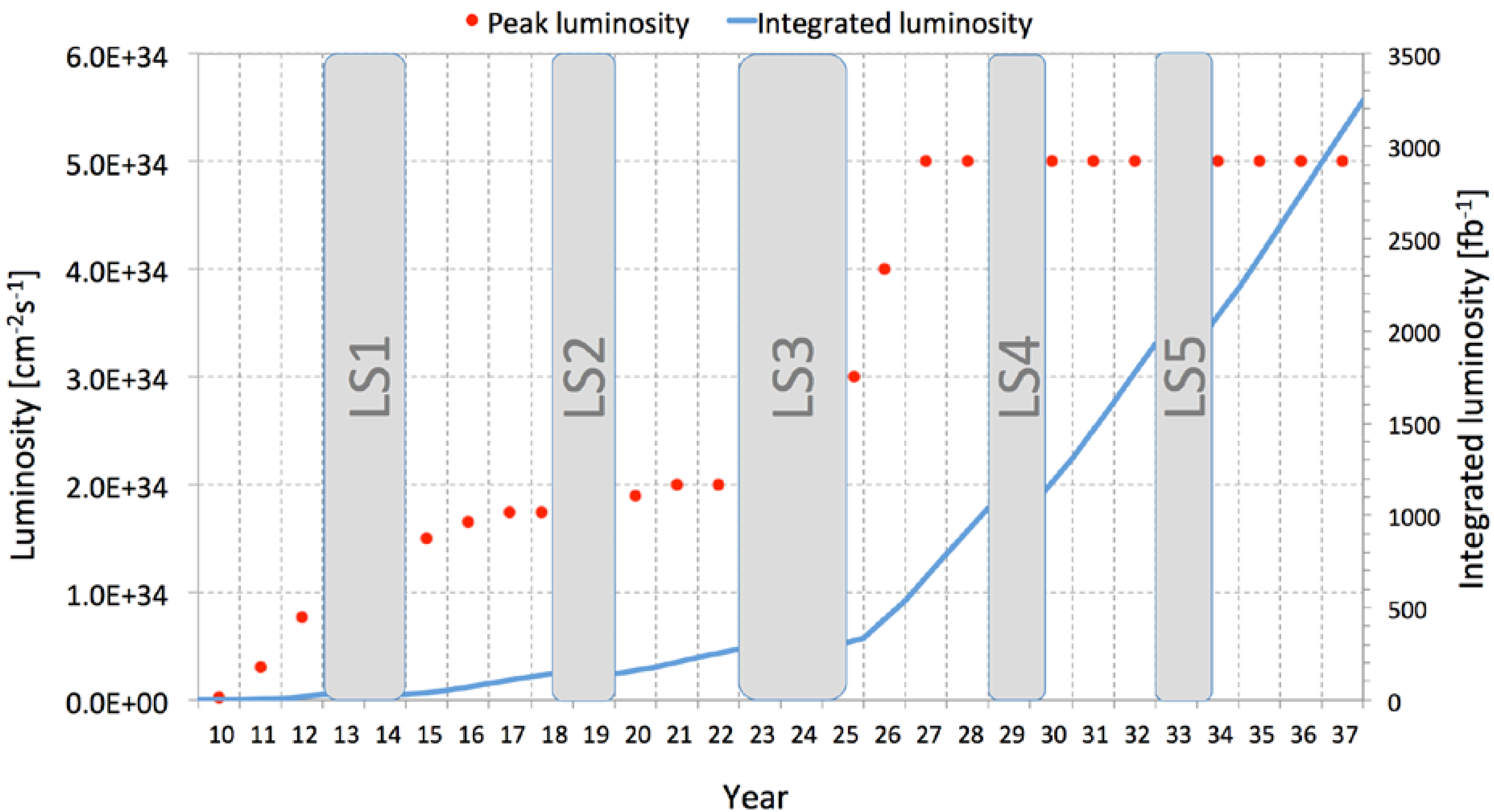
Moving forward, we still expect significant increase in mass reach.



Buckley, Feld, Macaluso, Monteux & DS 1610.08059

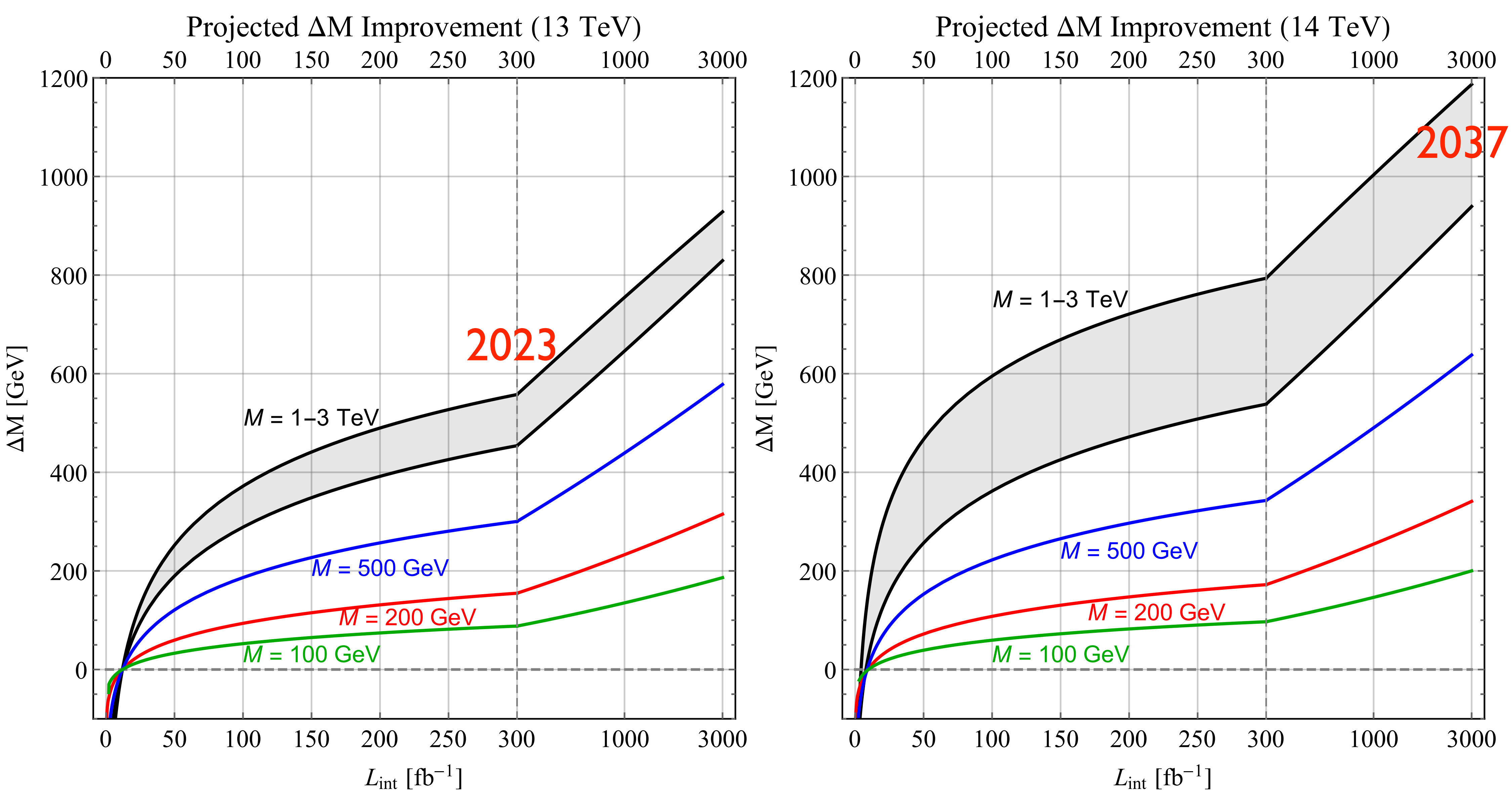
Assumptions: background, signal efficiencies unchanged, cross section controlled by parton luminosity divided by m^2. (cf. Salam & Weiler http://collider-reach.web.cern.ch/collider-reach/)

But progress will be much slower...



(M. Lamont, Moriond 2015)

Doubling the mass reach on ~TeV sparticles will take another 20 years!!



Buckley, Feld, Macaluso, Monteux & DS 1610.08059

Many people seem to be depressed/worried/panicked by the null results from ICHEP and Moriond, and by the impending slowdown of progress. Many seem to think that the discovery potential of the LHC has already dried up.

I want to push back on this negative outlook in this talk. I believe it is premature and is based on a superficial reading of the data. There is a lot more information to be gleaned from digging deeper into the CMS and ATLAS searches.

There are now hundreds of signal regions in the CMS and ATLAS searches for new physics. The official propaganda plots of just a handful of simplified models explore only a small fraction of the SRs. They pass over many potentially interesting fluctuations in the data!

Even if there is no new physics, there will inevitably be some 2 and even 3sigma local fluctuations. In the past, a “wait and see” approach made a lot of sense, but as the data comes in more slowly, it becomes increasingly interesting to ask whether a collection of fluctuations can be fit by a model.

Some obvious benefits of playing this game:

- Reveal patterns of correlated fluctuations
- Provide a new target for search re-optimization
- Suggest new final states to search in
- Maybe one of the excesses will turn out to be real!

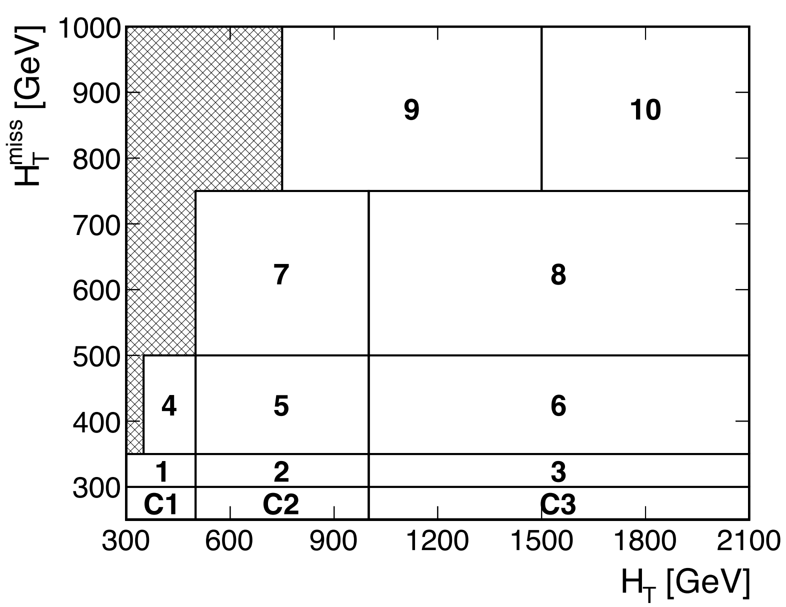
My collaborators and I are beginning to scratch the surface of this, starting with the jets+MET searches. We are finding a number of interesting excesses that have so far been overlooked!

work in progress with
Pouya Asadi, Matt Buckley, Anthony DiFranzo, Angelo Monteux...

CMS jets+MET searches

At Moriond, CMS presented two separate jets+MET searches. They are very similar. Both bin in HT, Nj and Nb. Main difference: choice of MET variable.

jets+MHT (CMS-16-033)



- N_{jet}: 2, 3–4, 5–6, 7–8, $\geq$9;
- $N_{b\text{-jet}}$: 0, 1, 2, $\geq$3.

jets+MT2 (CMS-16-036)

H_T range [GeV]	Jet multiplicities	M_{T2} binning [GeV]
[250, 450]	$2-3$j, 0b	[200, 300, 400, ∞)
	$2-3$j, 1b	[200, 300, 400, ∞)
	$2-3$j, 2b	[200, 300, 400, ∞)
	≥ 4j, 0b	[200, 300, 400, ∞)
	≥ 4j, 1b	[200, 300, 400, ∞)
	≥ 4j, 2b	[200, 300, 400, ∞)
	≥ 2j, ≥ 3b	[200, 300, 400, ∞)
[450, 575]	$2-3$j, 0b	[200, 300, 400, 500, ∞)
	$2-3$j, 1b	[200, 300, 400, 500, ∞)
	$2-3$j, 2b	[200, 300, 400, 500, ∞)
	$4-6$j, 0b	[200, 300, 400, 500, ∞)
	$4-6$j, 1b	[200, 300, 400, 500, ∞)
	$4-6$j, 2b	[200, 300, 400, 500, ∞)
	≥ 7j, 0b	[200, 300, 400, ∞)
	≥ 7j, 1b	[200, 300, 400, ∞)
	≥ 7j, 2b	[200, 300, 400, ∞)
	$2-6$j, ≥ 3b	[200, 300, 400, 500, ∞)
	≥ 7j, ≥ 3b	[200, 300, 400, ∞)
[575, 1000]	$2-3$j, 0b	[200, 300, 400, 600, 800, ∞)
	$2-3$j, 1b	[200, 300, 400, 600, 800, ∞)
	$2-3$j, 2b	[200, 300, 400, 600, 800, ∞)
	$4-6$j, 0b	[200, 300, 400, 600, 800, ∞)
	$4-6$j, 1b	[200, 300, 400, 600, 800, ∞)
	$4-6$j, 2b	[200, 300, 400, 600, 800, ∞)
	≥ 7j, 0b	[200, 300, 400, 600, 800, ∞)
	≥ 7j, 1b	[200, 300, 400, 600, ∞)
	≥ 7j, 2b	[200, 300, 400, 600, ∞)
	$2-6$j, ≥ 3b	[200, 300, 400, 600, ∞)
	≥ 7j, ≥ 3b	[200, 300, 400, 600, ∞)

H_T range [GeV]	Jet multiplicities	M_{T2} binning [GeV]
[1000, 1500]	$2-3$j, 0b	[200, 400, 600, 800, 1000, 1200, ∞)
	$2-3$j, 1b	[200, 400, 600, 800, 1000, 1200, ∞)
	$2-3$j, 2b	[200, 400, 600, 800, 1000, ∞)
	$4-6$j, 0b	[200, 400, 600, 800, 1000, 1200, ∞)
	$4-6$j, 1b	[200, 400, 600, 800, 1000, 1200, ∞)
	$4-6$j, 2b	[200, 400, 600, 800, 1000, ∞)
	≥ 7j, 0b	[200, 400, 600, 800, 1000, ∞)
	≥ 7j, 1b	[200, 400, 600, 800, ∞)
	≥ 7j, 2b	[200, 400, 600, 800, ∞)
	$2-6$j, ≥ 3b	[200, 400, 600, ∞)
	≥ 7j, ≥ 3b	[200, 400, 600, ∞)
[1500, ∞)	$2-3$j, 0b	[400, 600, 800, 1000, 1400, ∞)
	$2-3$j, 1b	[400, 600, 800, 1000, ∞)
	$2-3$j, 2b	[400, ∞)
	$4-6$j, 0b	[400, 600, 800, 1000, 1400, ∞)
	$4-6$j, 1b	[400, 600, 800, 1000, 1400, ∞)
	$4-6$j, 2b	[400, 600, 800, ∞)
	≥ 7j, 0b	[400, 600, 800, 1000, ∞)
	≥ 7j, 1b	[400, 600, 800, ∞)
	≥ 7j, 2b	[400, 600, 800, ∞)
	$2-6$j, ≥ 3b	[400, 600, ∞)
	≥ 7j, ≥ 3b	[400, ∞)

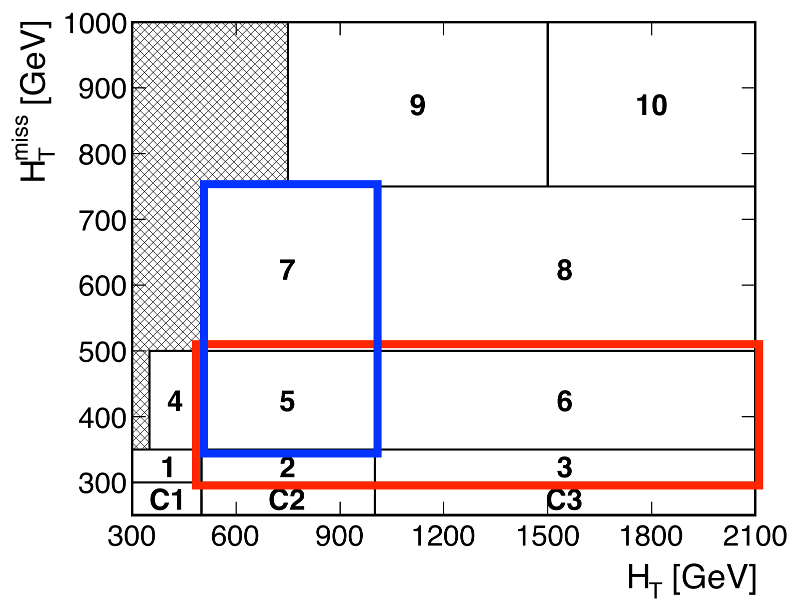
$$M_{T2} = \min_{\vec{p}_T^{\,\text{miss}X(1)} + \vec{p}_T^{\,\text{miss}X(2)} = \vec{p}_T^{\,\text{miss}}} \left[\max\left(M_T^{(1)}, M_T^{(2)} \right) \right]$$

Problem: too many signal regions

Each CMS search generally consists of hundreds of SRs defined by binning in several kinematic variables. But these SRs are sliced so finely, an excess from any realistic signal would likely span multiple neighboring SRs.

Idea: consider all possible "rectangular aggregations" as a way of enhancing signal vs background.

jets+MHT (CMS-16-033)



- N_{jet}: 2, 3–4, 5–6, 7–8, ≥ 9;
- $N_{\text{b-jet}}$: 0, 1, 2, ≥ 3.

Mining rectangular aggregations

Plan: make a list of the most significant excesses, test for compatibility with other searches, try to find models that fit them.

To avoid overcounting, discard RAs that contain other nearly-as-significant aggregations.

Results:

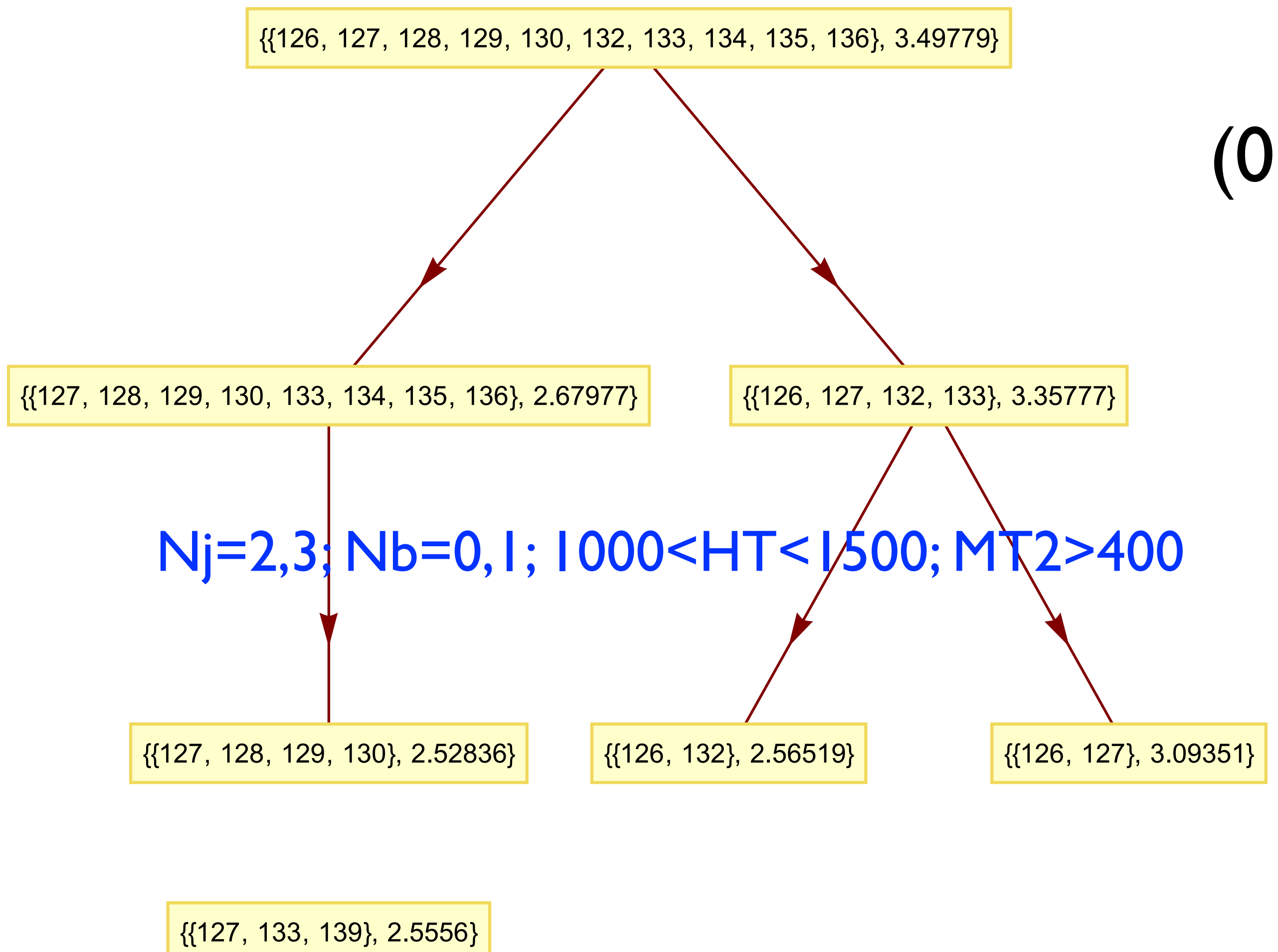
033 jets+MHT

- ~7000 rectangular aggregations
- 13 with ≥ 2.5 sigma

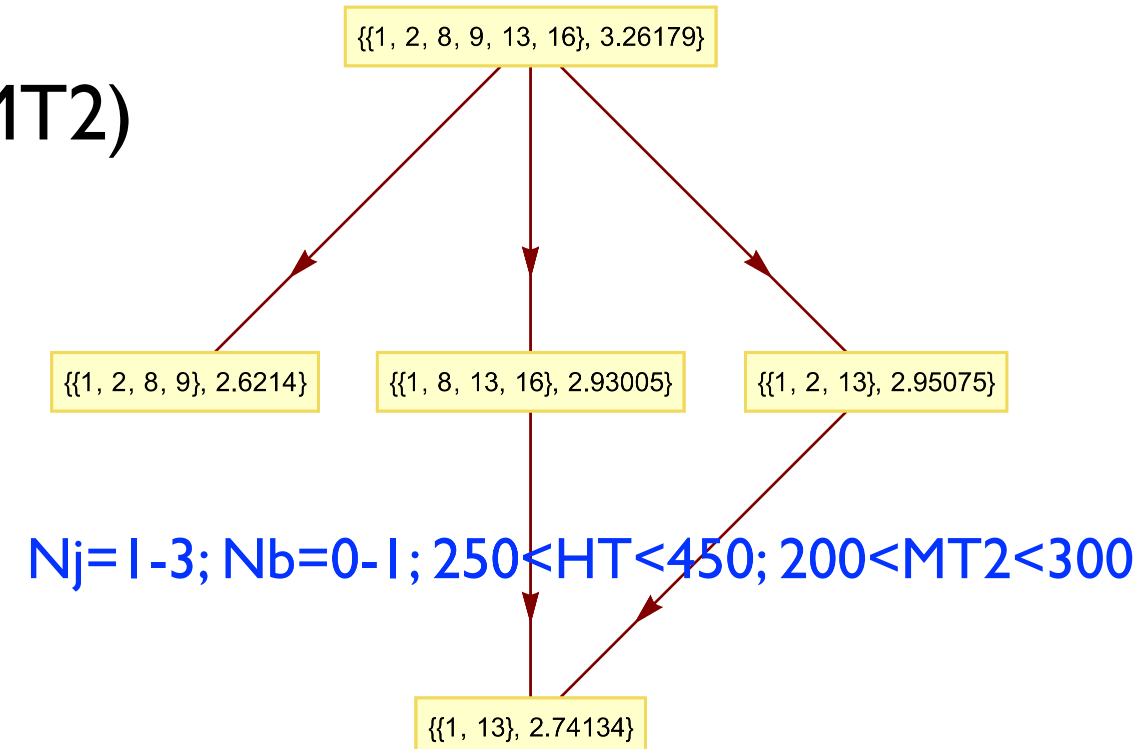
036 jets+MT2

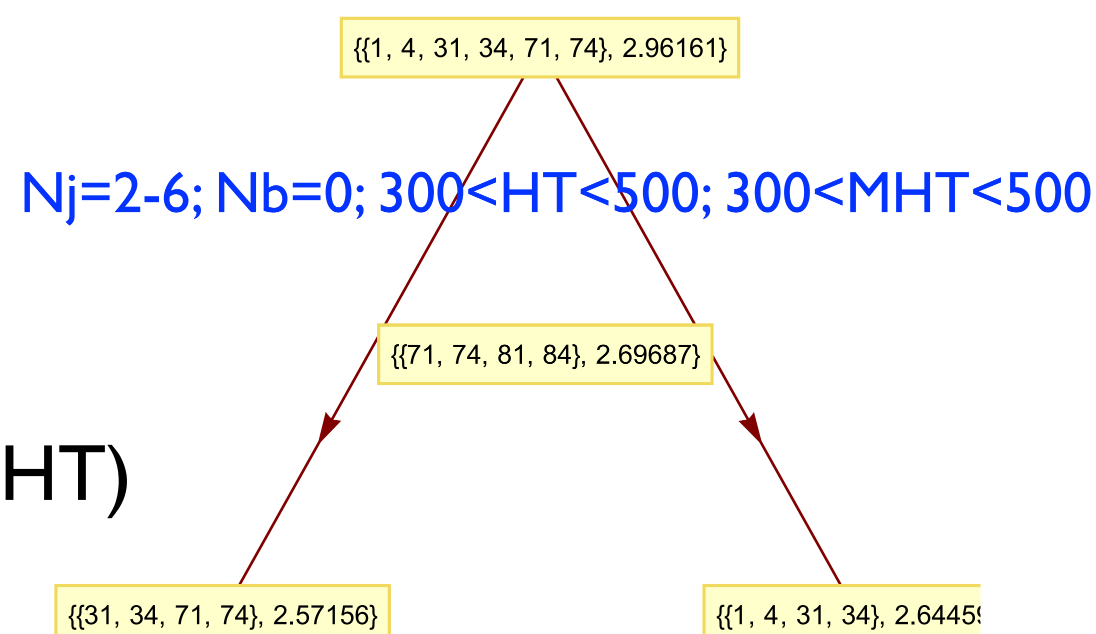
- ~33000 rectangular aggregations
- 17 with ≥ 2.5 sigma

These excesses break up further into a handful of clusters. Here are a few of the most significant ones (there are several more):



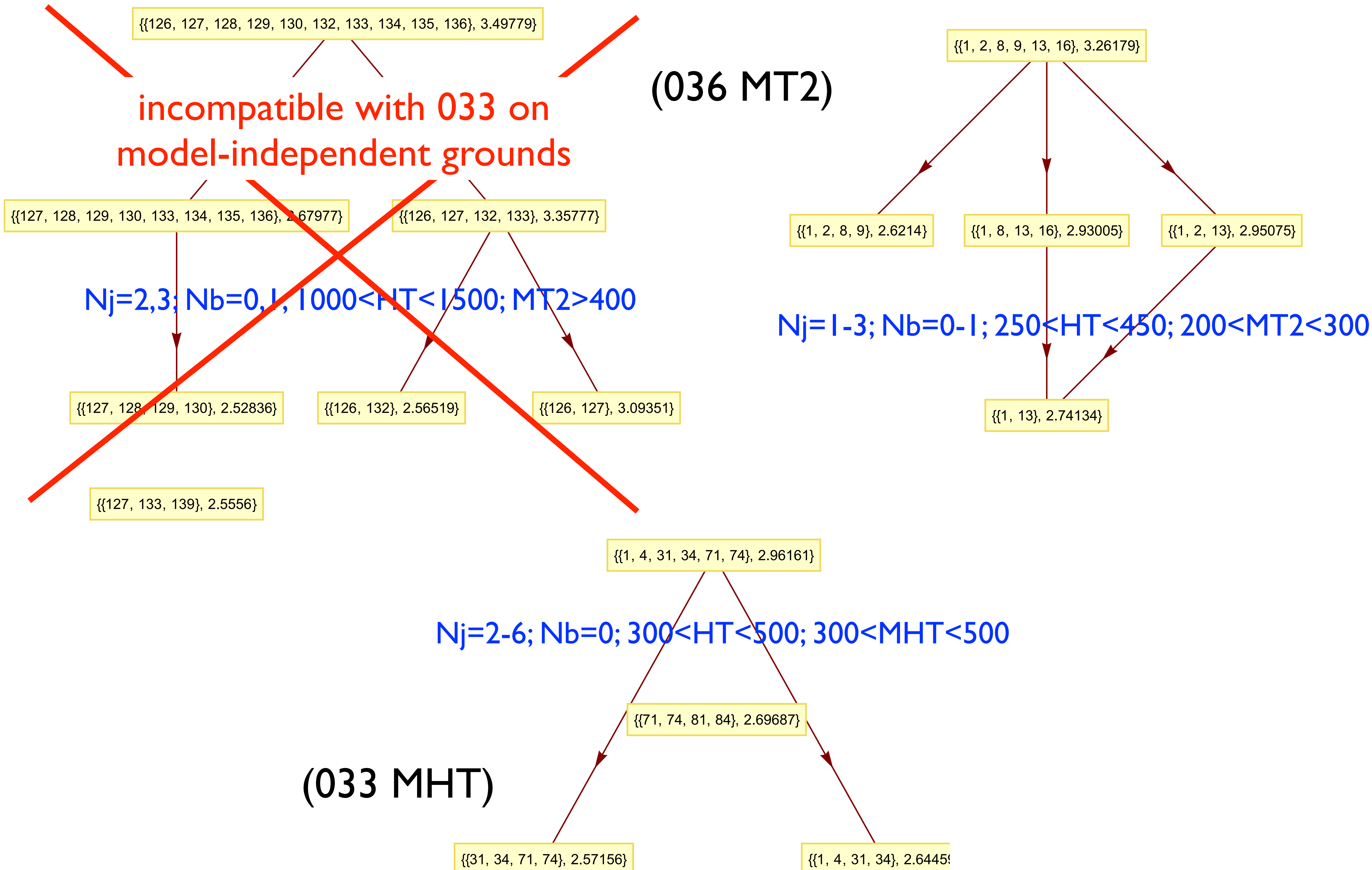
(036 MT2)



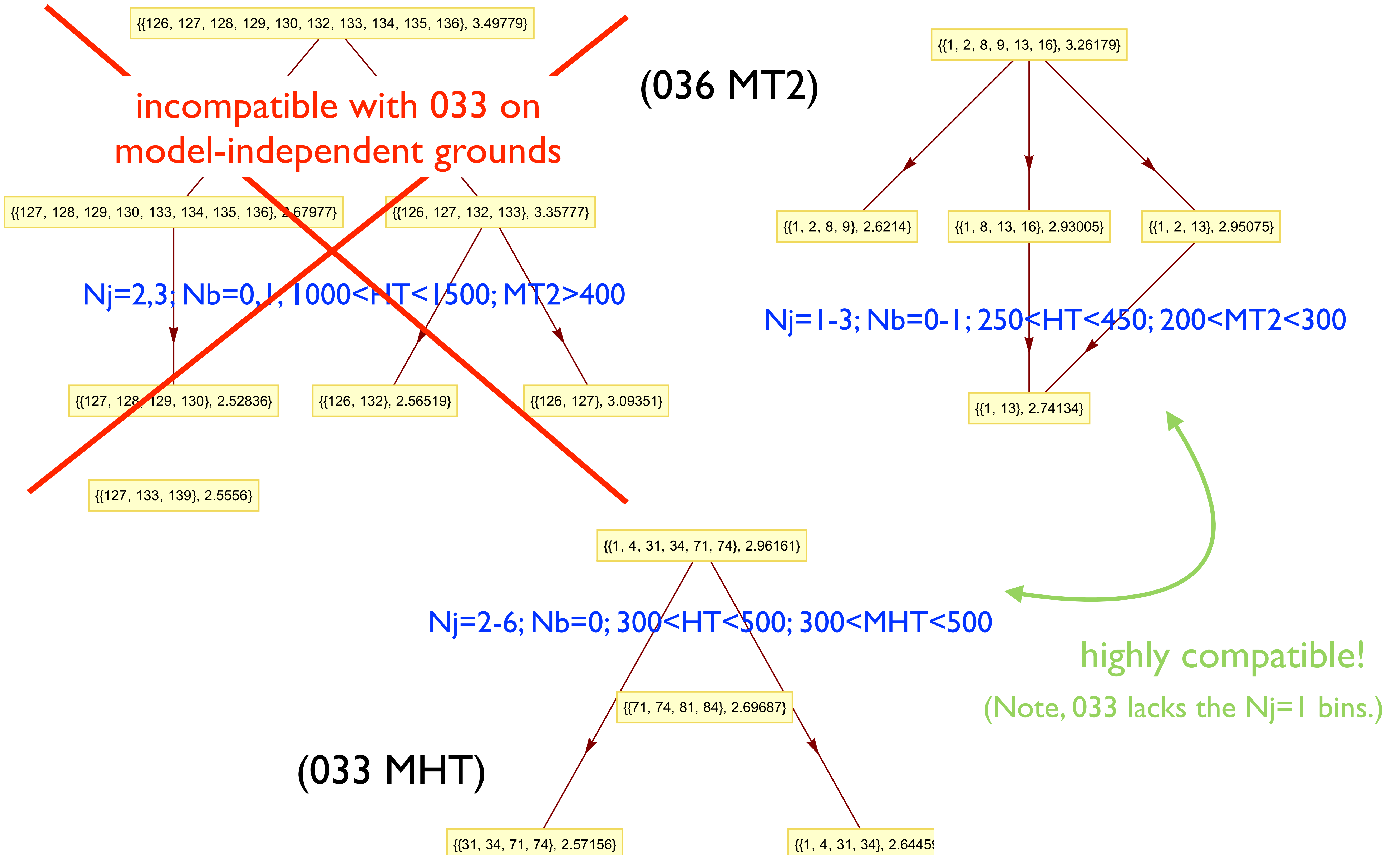


(033 MHT)

These excesses break up further into a handful of clusters. Here are a few of the most significant ones (there are several more):



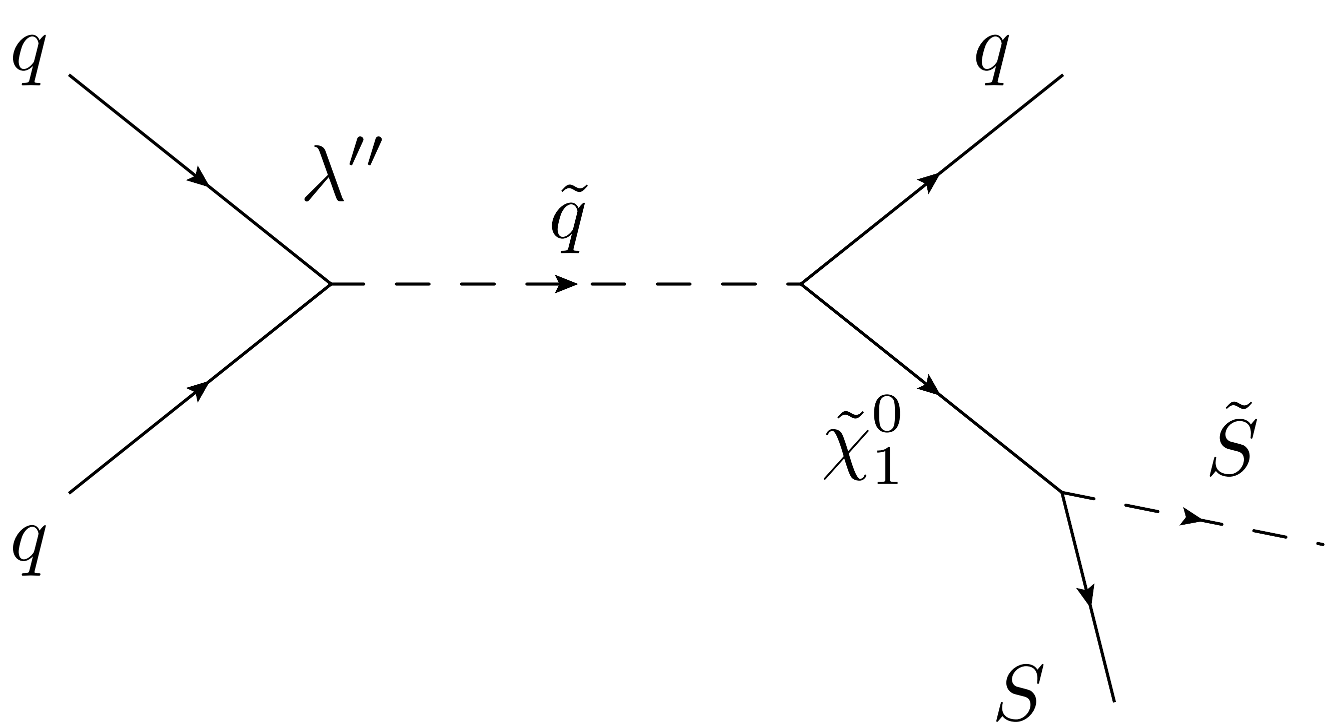
These excesses break up further into a handful of clusters. Here are a few of the most significant ones (there are several more):



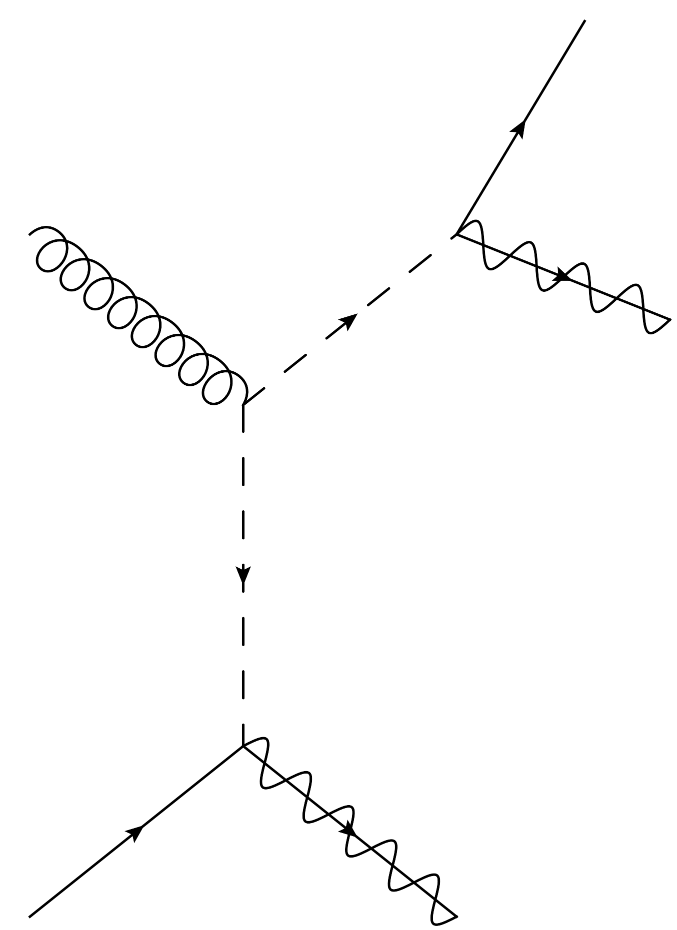
Fitting to a model -- two candidates

Excess is driven largely by Nj=1-3, Nb=0, 250<HT<450; 200<MT2<300

Idea: one parton + invisible particle(s) in the hard process, rest from ISR/FSR



“monosquark” model



squark-neutralino associated production

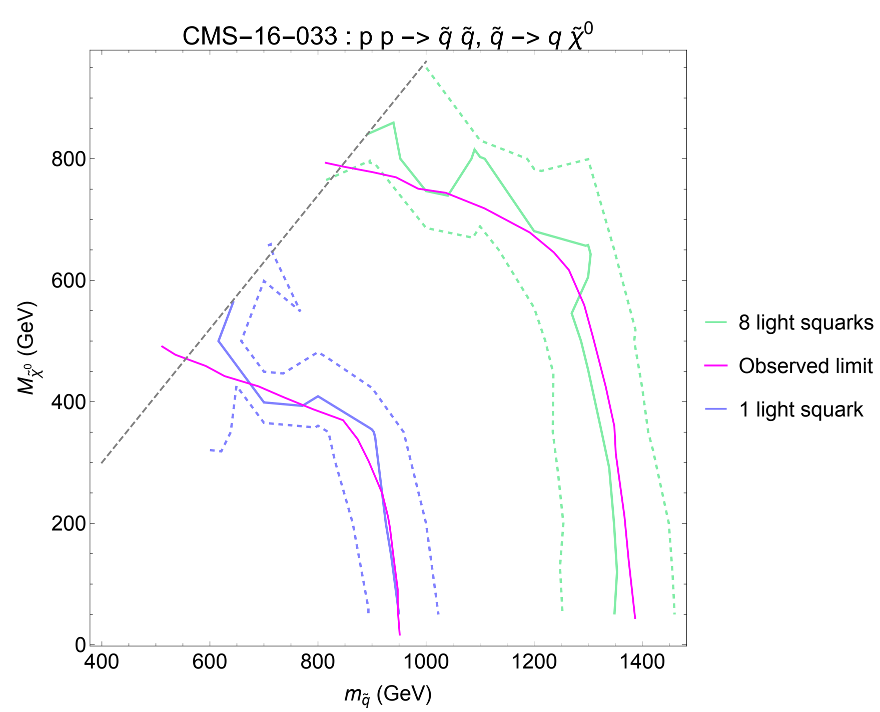
Our recasting pipeline

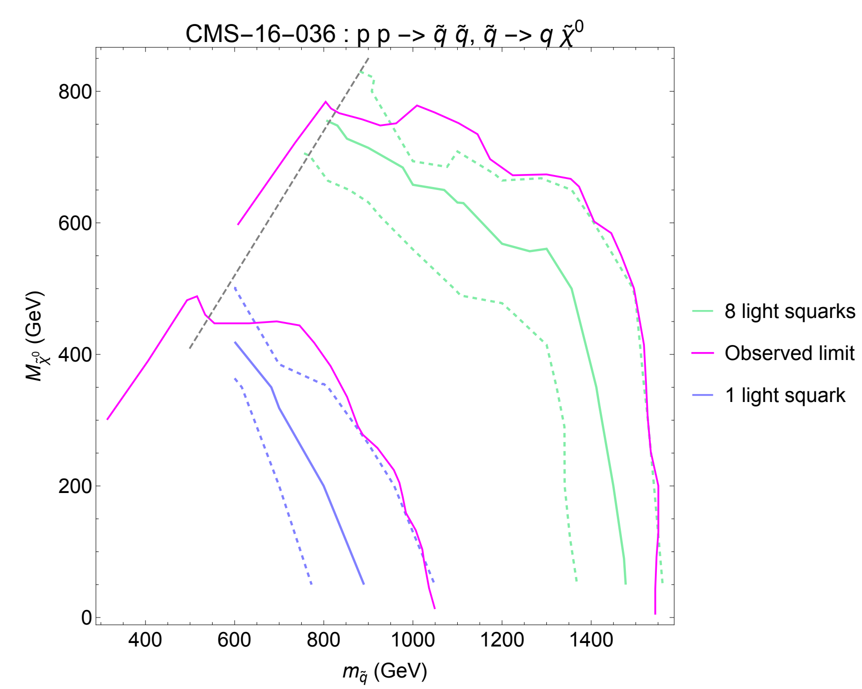
Buckley, Feld, Macaluso, Monteux & DS 1610.08059

Madgraph5; Prospino 2.1 — **Hard process (signal only)** → Pythia8.2 — **Showering, hadronization** → Delphes3 — **Detector simulation**

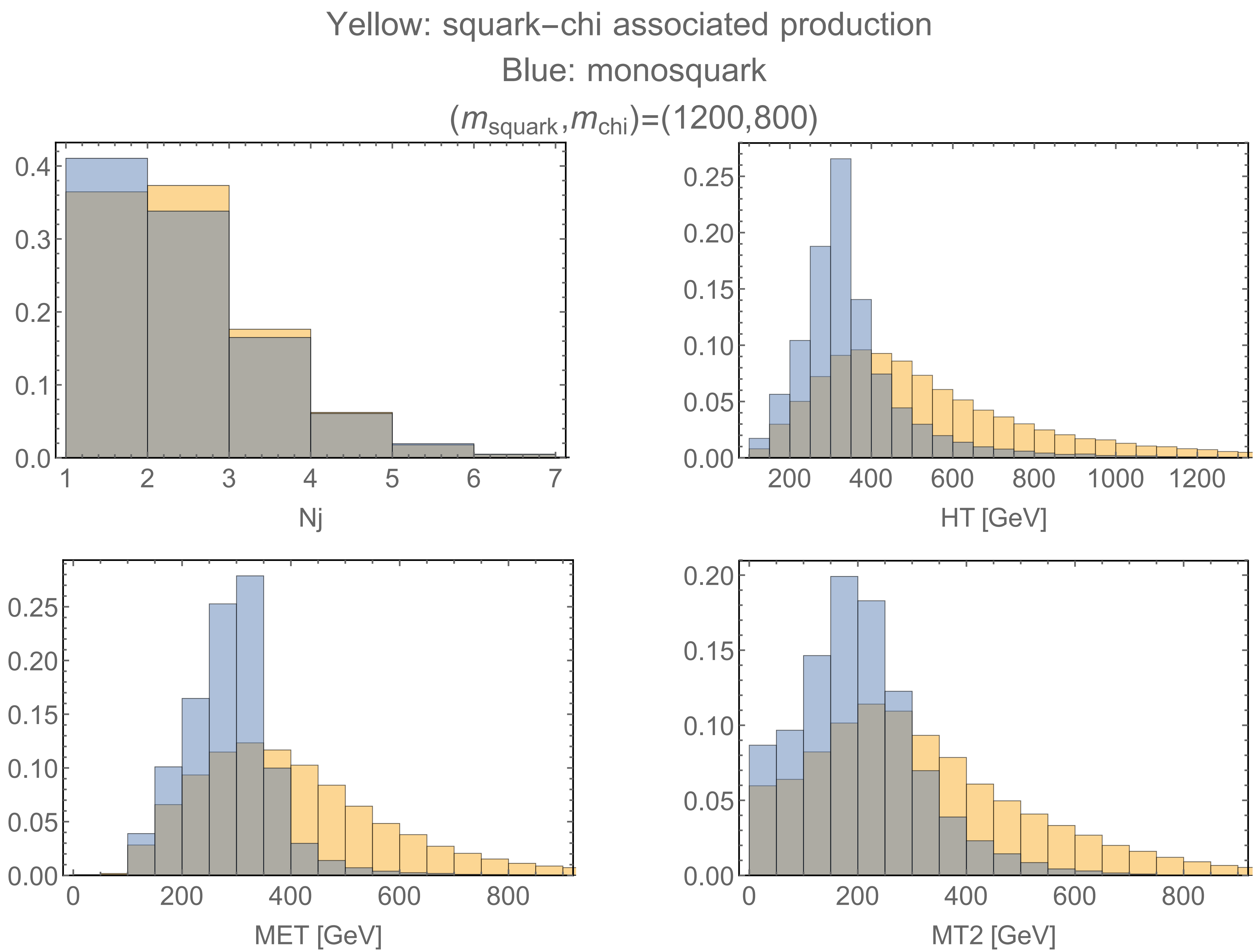
↓

Limit plots ← **Signal efficiencies** ← **Recasted ATLAS/CMS analysis**





Nj=1-3; Nb=0; 250<HT<450; 200<MT2<300



Monosquark model looks promising!

Lightning stats review

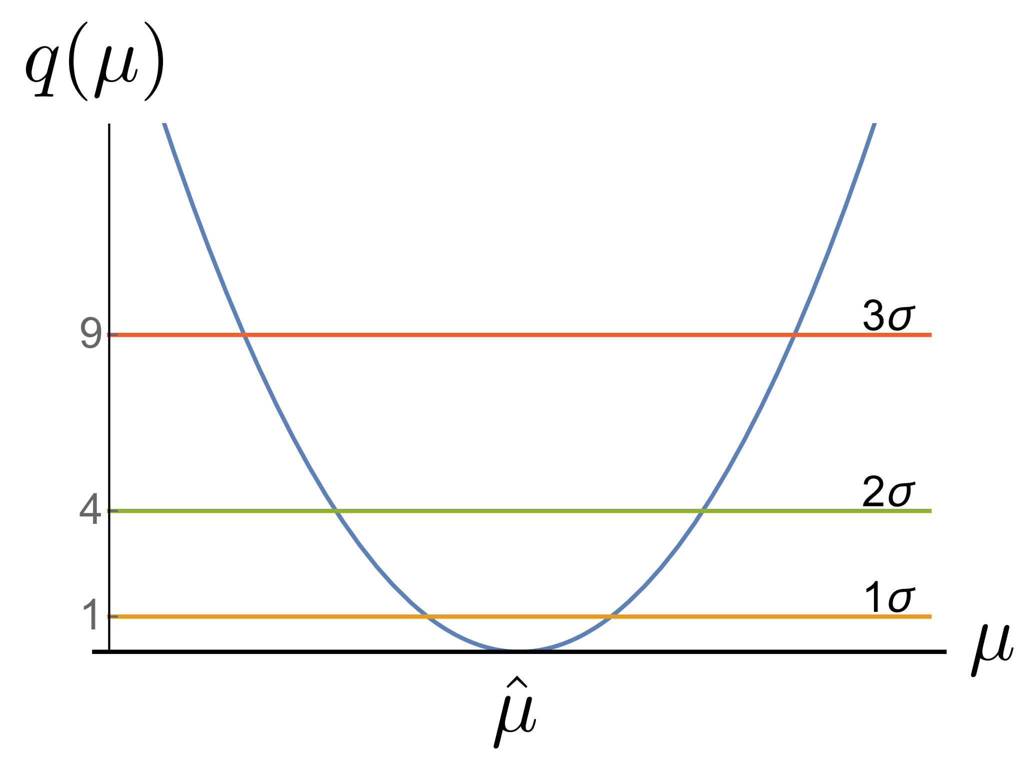
We use the asymptotic profile likelihood methods of Cowan, Cranmer, Gross & Vitells (1007.1727). We take into account correlations using the covariance matrices provided by CMS. (The correlations are essential!!)

$$L(n,p,\theta) = \prod_i \frac{1}{n_i!}(p_i+\theta_i)^{n_i} e^{-(p_i+\theta_i)} e^{-\frac{1}{2}\theta^T C^{-1}\theta}$$

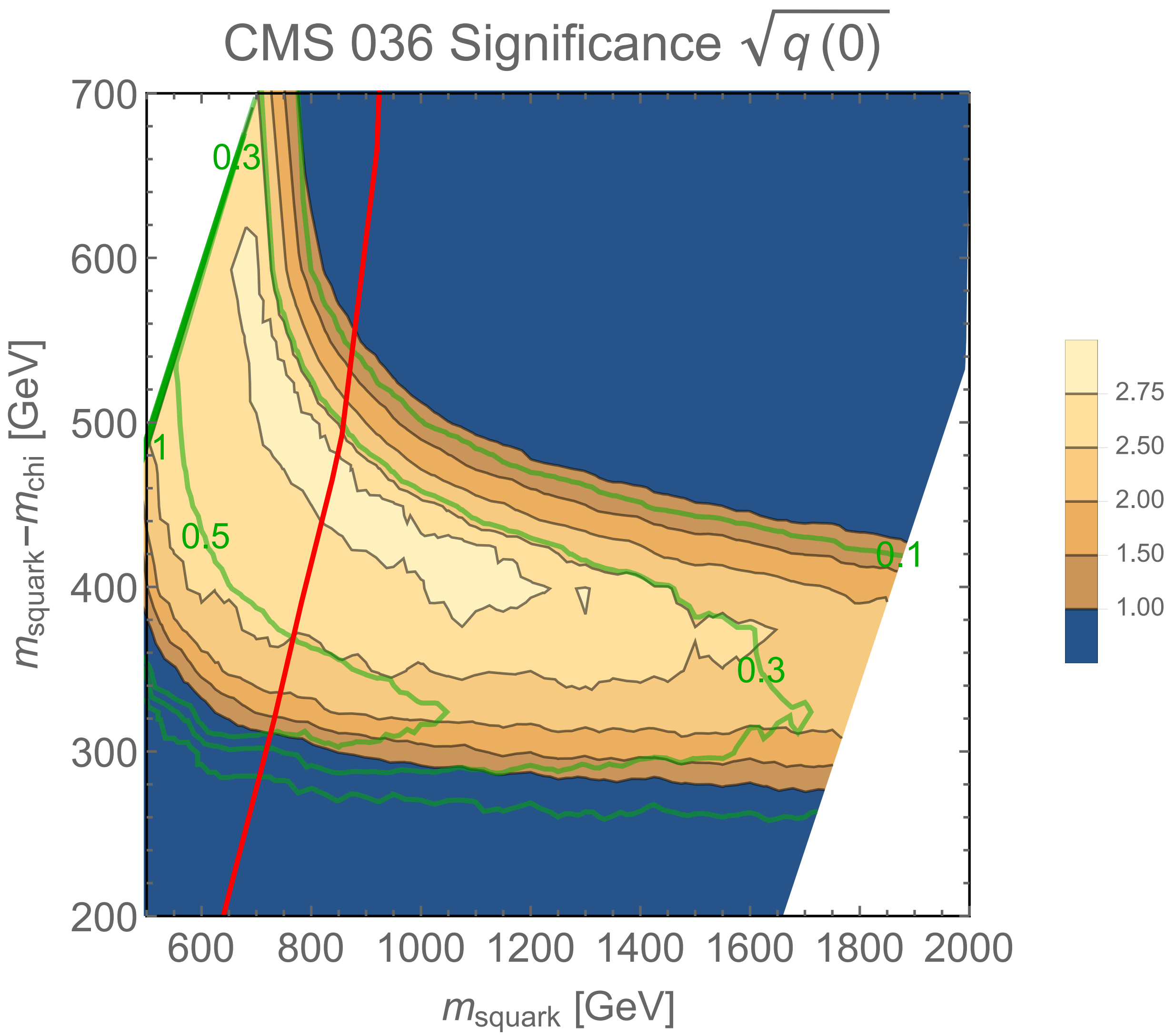
$$L(\mu,\theta) = L(n_i, p_i + \mu s_i, \theta_i)$$

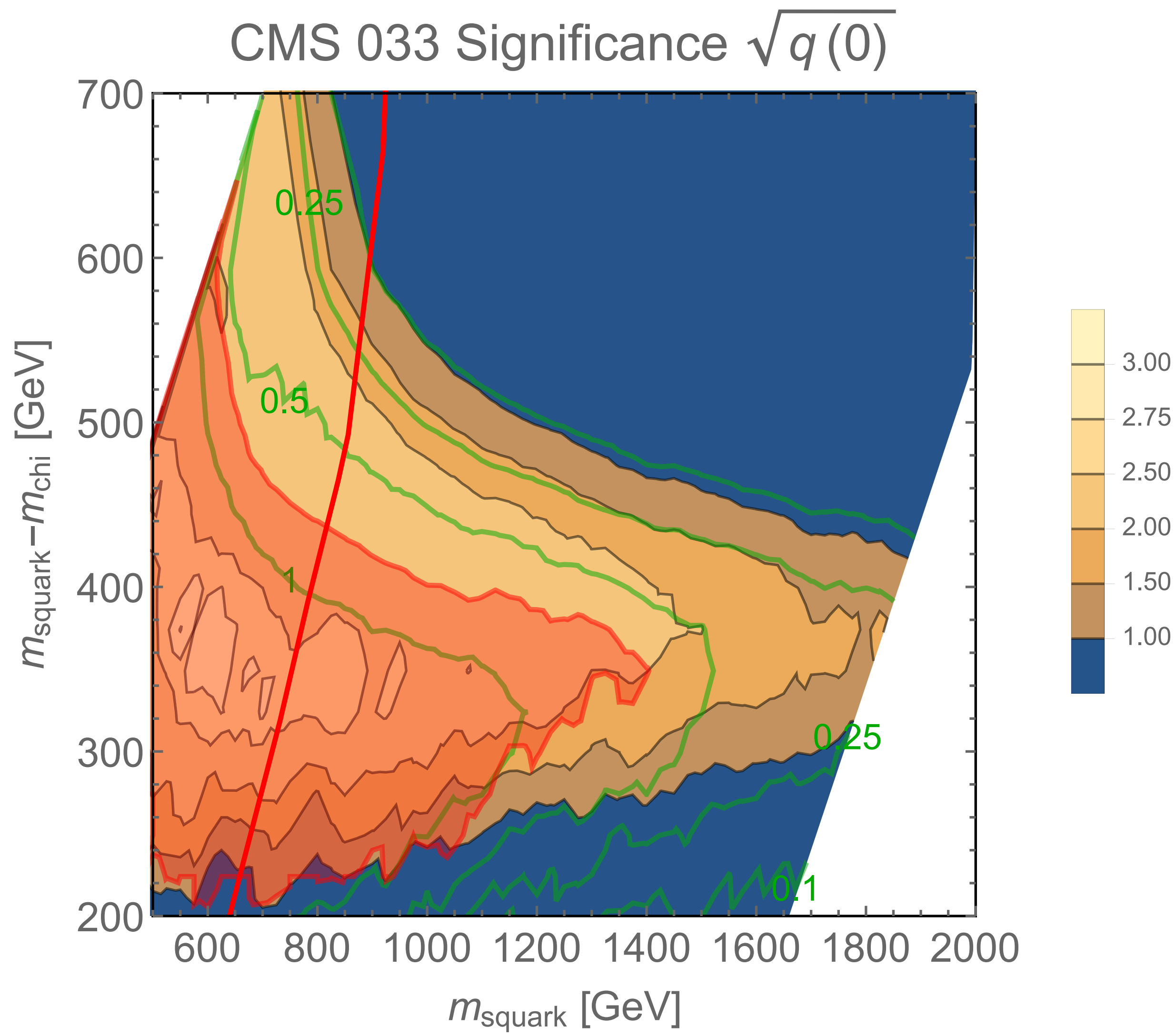
$$\lambda(\mu) = \frac{L(\mu,\hat{\theta}(\mu))}{L(\hat{\mu},\hat{\theta}(\hat{\mu}))}$$

$$q(\mu) = -2\log\lambda(\mu)$$



Significance plots

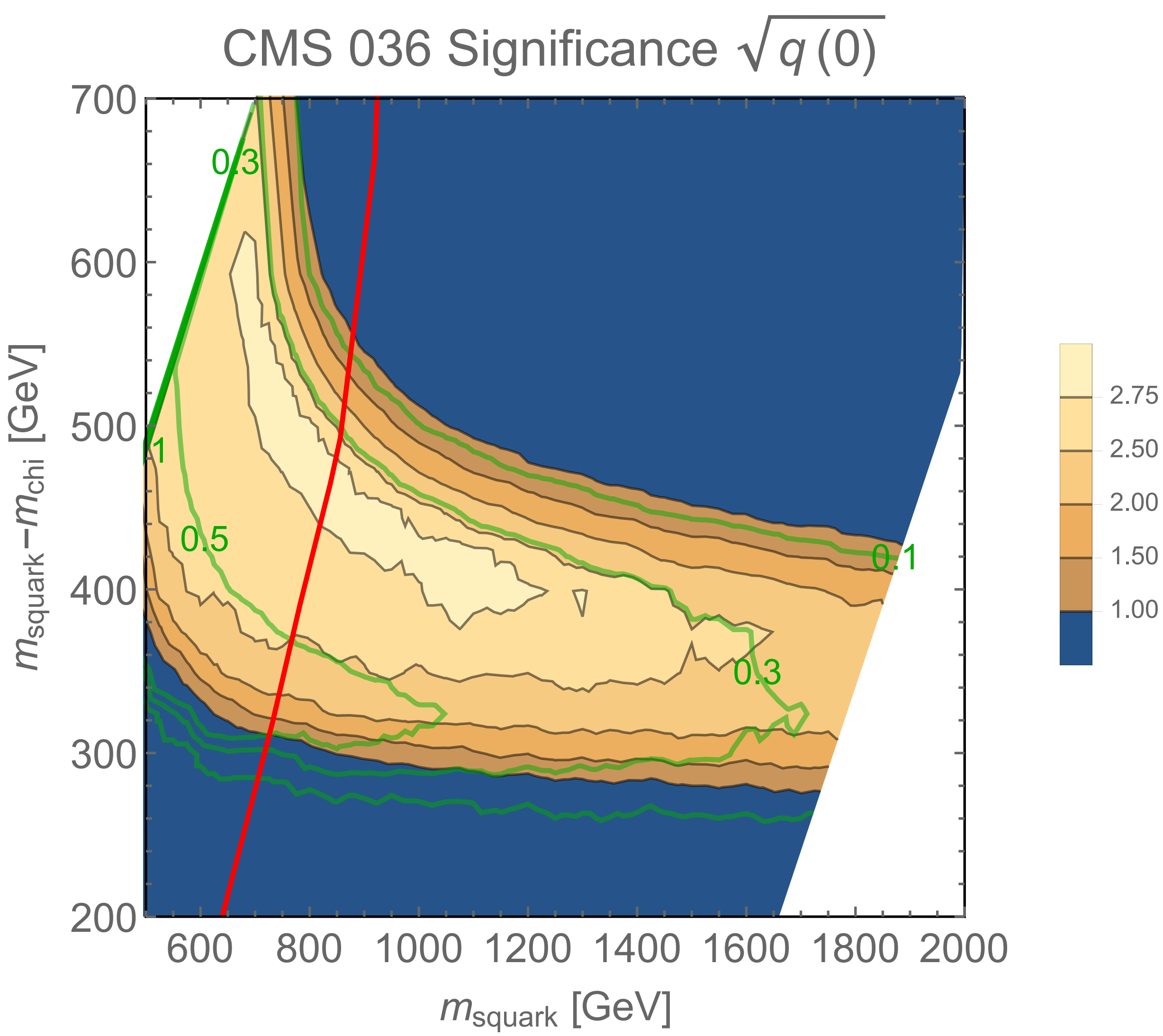


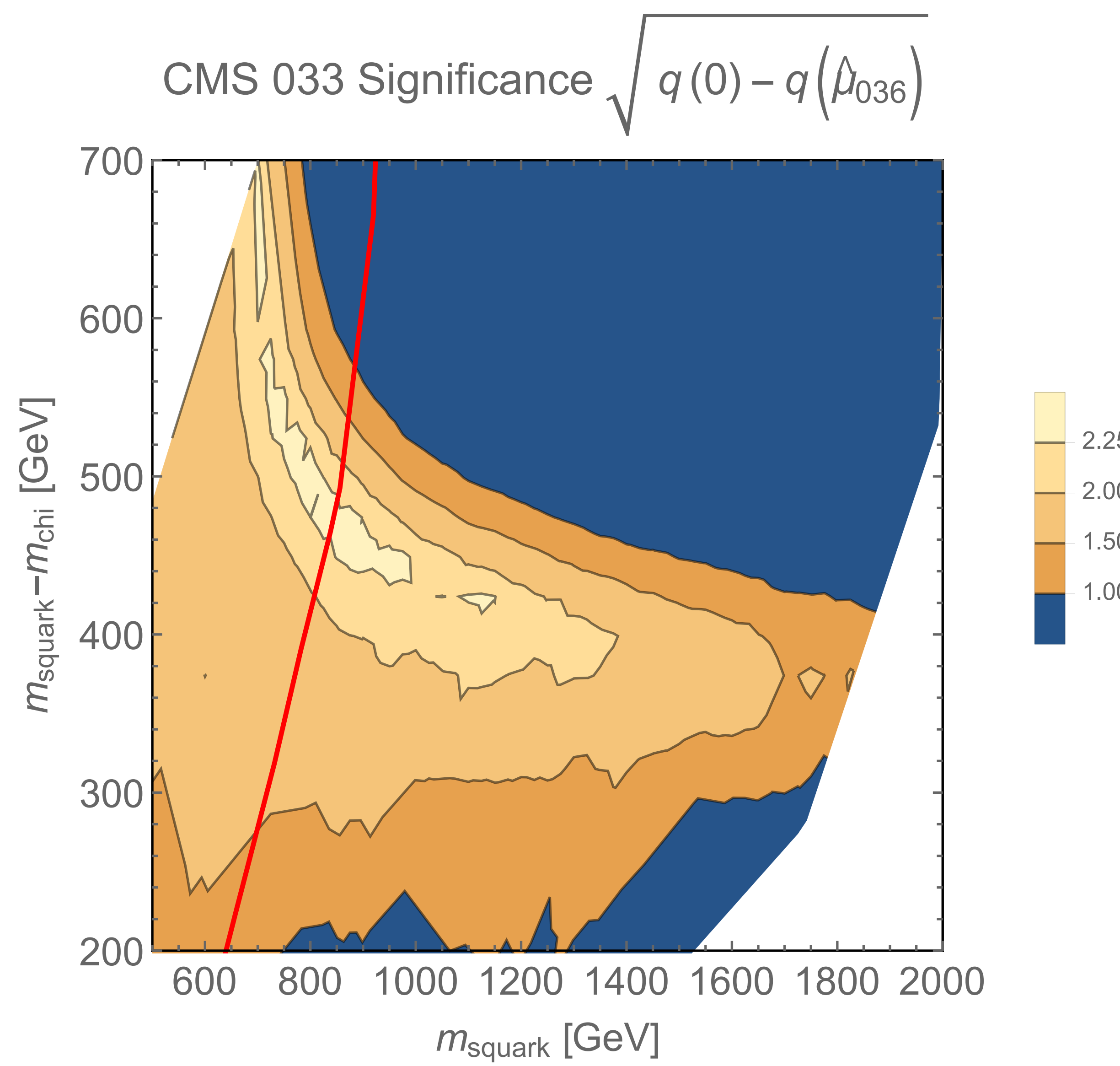


Best fit cross section for 036 is entirely compatible with 033 and can reach nearly 3 sigma local significance!

Best fit cross section for 033 is larger, in tension with 036.

Significance plots





At best fit signal strength for 036, can explain ~2 sigma of the 033 excess!

Adding in ATLAS

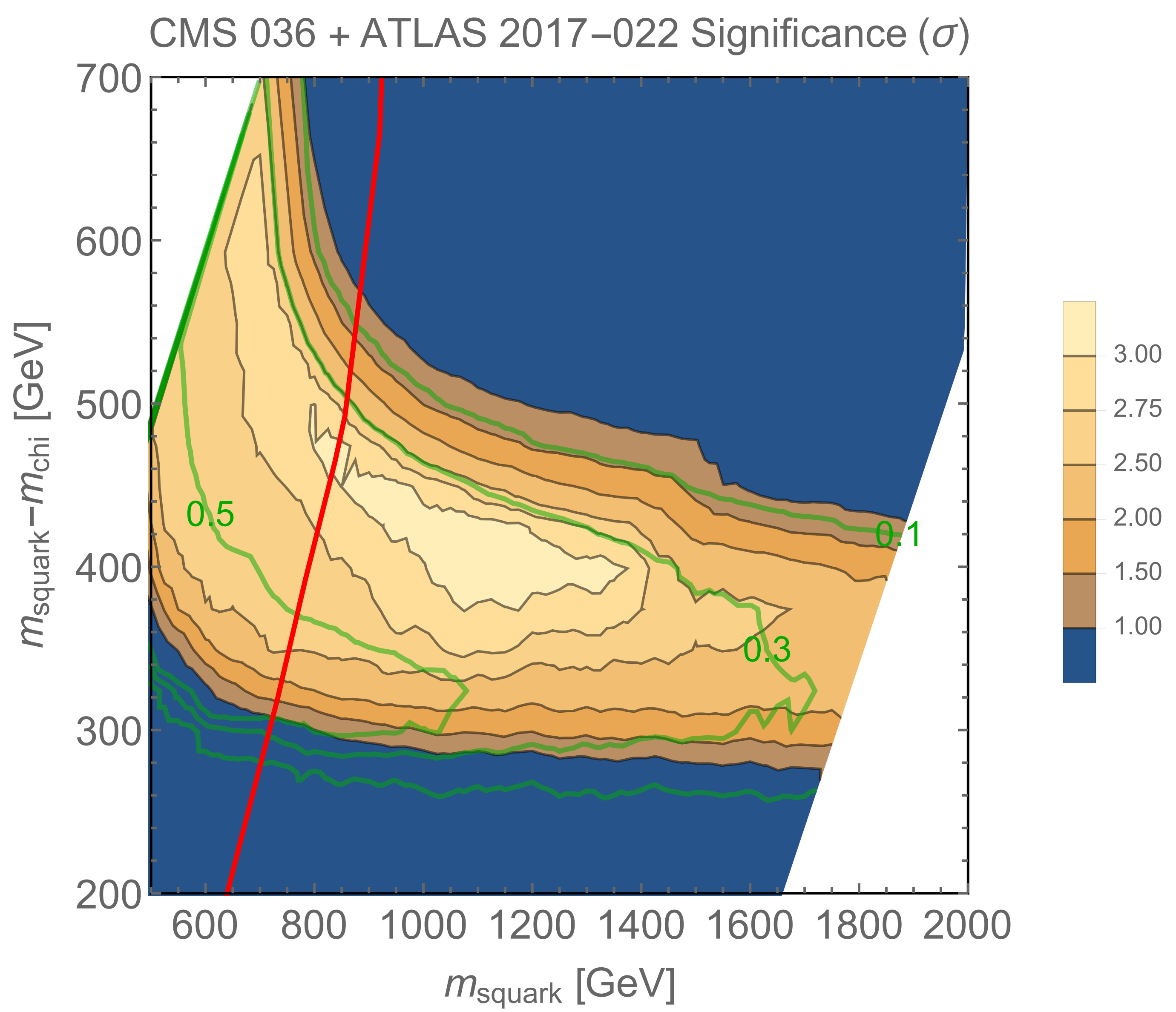
What about ATLAS? They also presented a jets+MET search at Moriond (2017-022). But they did not bin as finely as CMS, and they did not explore the low pT regions of their dataset. As a result, they are not very sensitive to this model.

Nevertheless, they did see a slight excess in their lowest Nj, lowest Meff SR...

Targeted signal	$\tilde{q}\tilde{q}$, $\tilde{q} \rightarrow q\tilde{\chi}_1^0$							
Requirement	Signal Region [**Meff-**]							
	2j-1200	**2j-1600**	**2j-2000**	**2j-2400**	**2j-2800**	**2j-3600**	**2j-2100**	**3j-1300**
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV] >	250							
$p_{\mathrm{T}}(j_1)$ [GeV] >	250	300	350				600	700
$p_{\mathrm{T}}(j_2)$ [GeV] >	250	300	350				50	
$p_{\mathrm{T}}(j_3)$ [GeV] >	–							50
$\lvert\eta(j_{1,2})\rvert$ <	0.8	1.2				–		
$\Delta\phi(\mathrm{jet}_{1,2,(3)}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}}$ >	0.8						0.4	
$\Delta\phi(\mathrm{jet}_{i>3}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}}$ >	0.4						0.2	
$E_{\mathrm{T}}^{\mathrm{miss}} / \sqrt{H_{\mathrm{T}}}$ [GeV$^{1/2}$] >	14	18					26	16
m_{eff}(incl.) [GeV] >	1200	1600	2000	2400	2800	3600	2100	1300

Signal Region [**Meff-**]	**2j-1200**
Diboson	28.17
Z/γ^*+jets	346.37
W+jets	142.39
$t\bar{t}$(+EW) + single top	21.40
Diboson	28 ± 4
Z/γ^*+jets	337 ± 19
W+jets	136 ± 24
$t\bar{t}$(+EW) + single top	15 ± 4
Multi-jet	1.8 ± 1.8
Total bkg	517 ± 31
Observed	582
$\langle\epsilon\sigma\rangle_{\mathrm{obs}}^{95}$ [fb]	3.6
S_{obs}^{95}	131
S_{exp}^{95}	78^{+33}_{-21}
p_0 (Z)	0.06 (1.53)

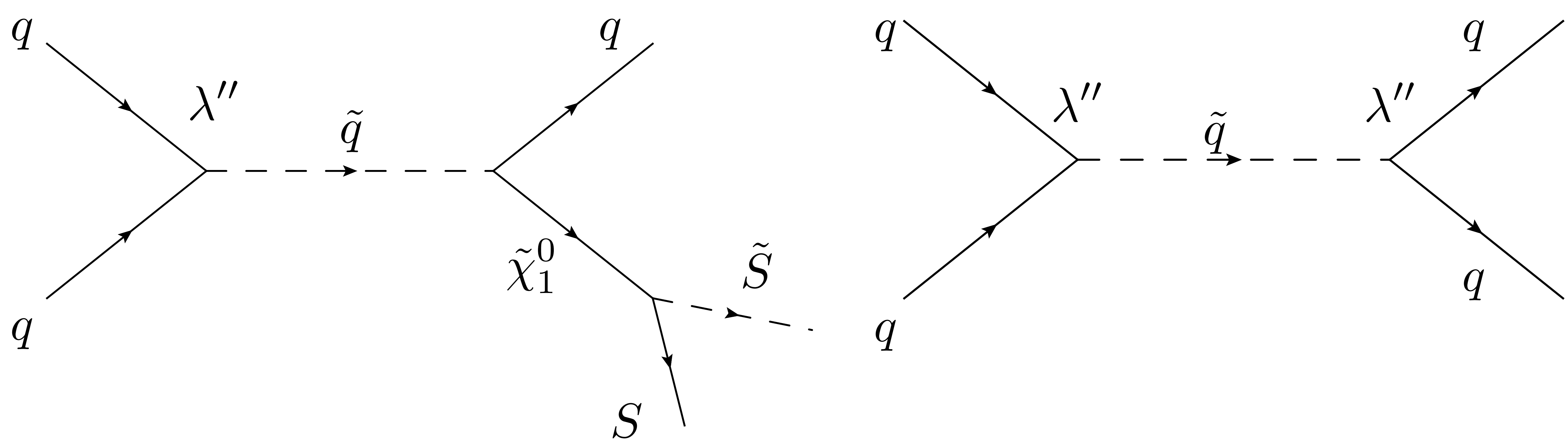
Adding in ATLAS



Adding in ATLAS improves the local significance to over 3sigma!

(Best point has 3.5sigma!)

Dijet resonances



Dijet resonances are an important correlated signature of the model.

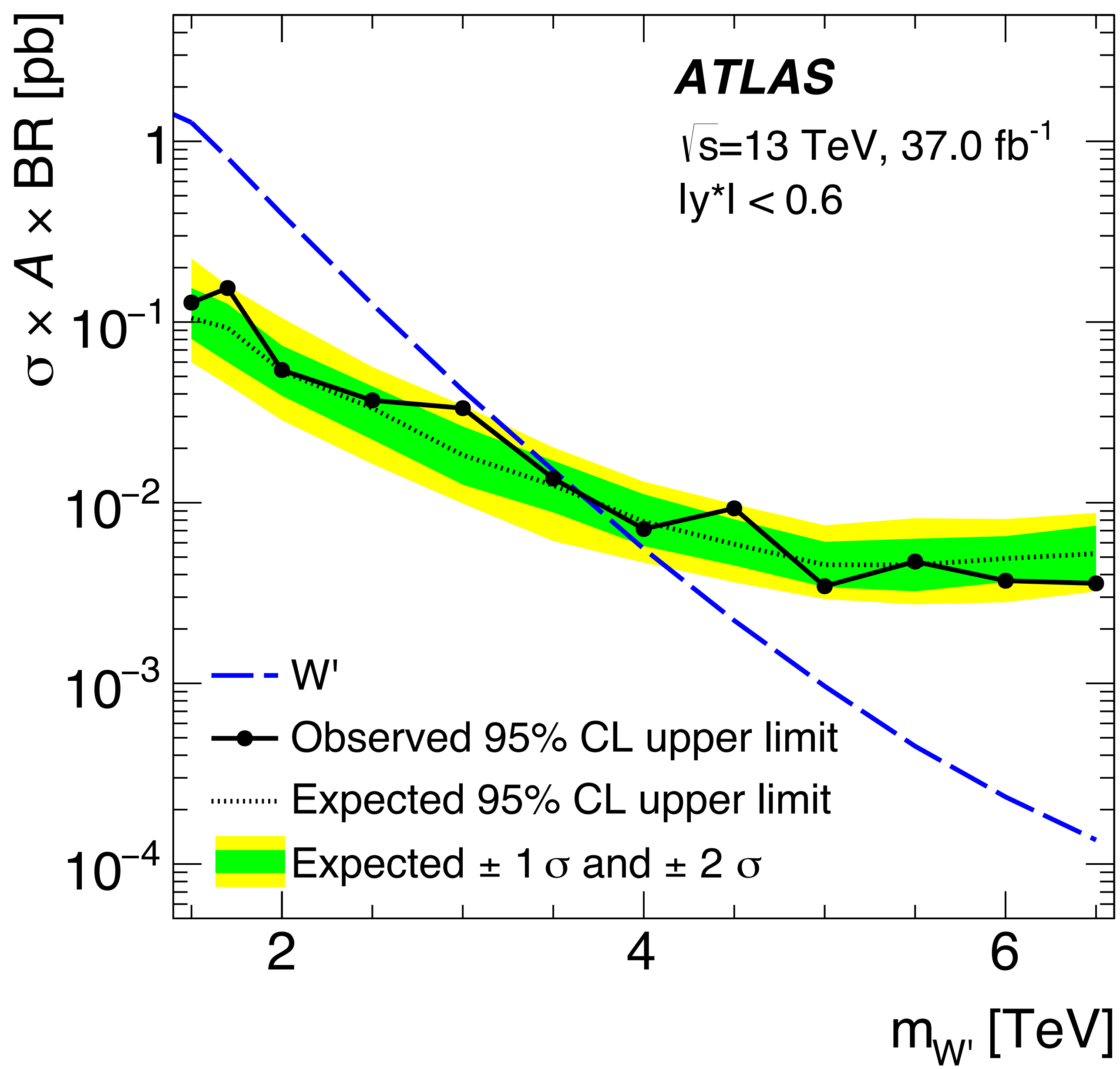
Details are model dependent...

Dijet resonances

ATLAS and CMS both presented dijet resonance searches at Moriond.

ATLAS doesn't go below 1.5 TeV.

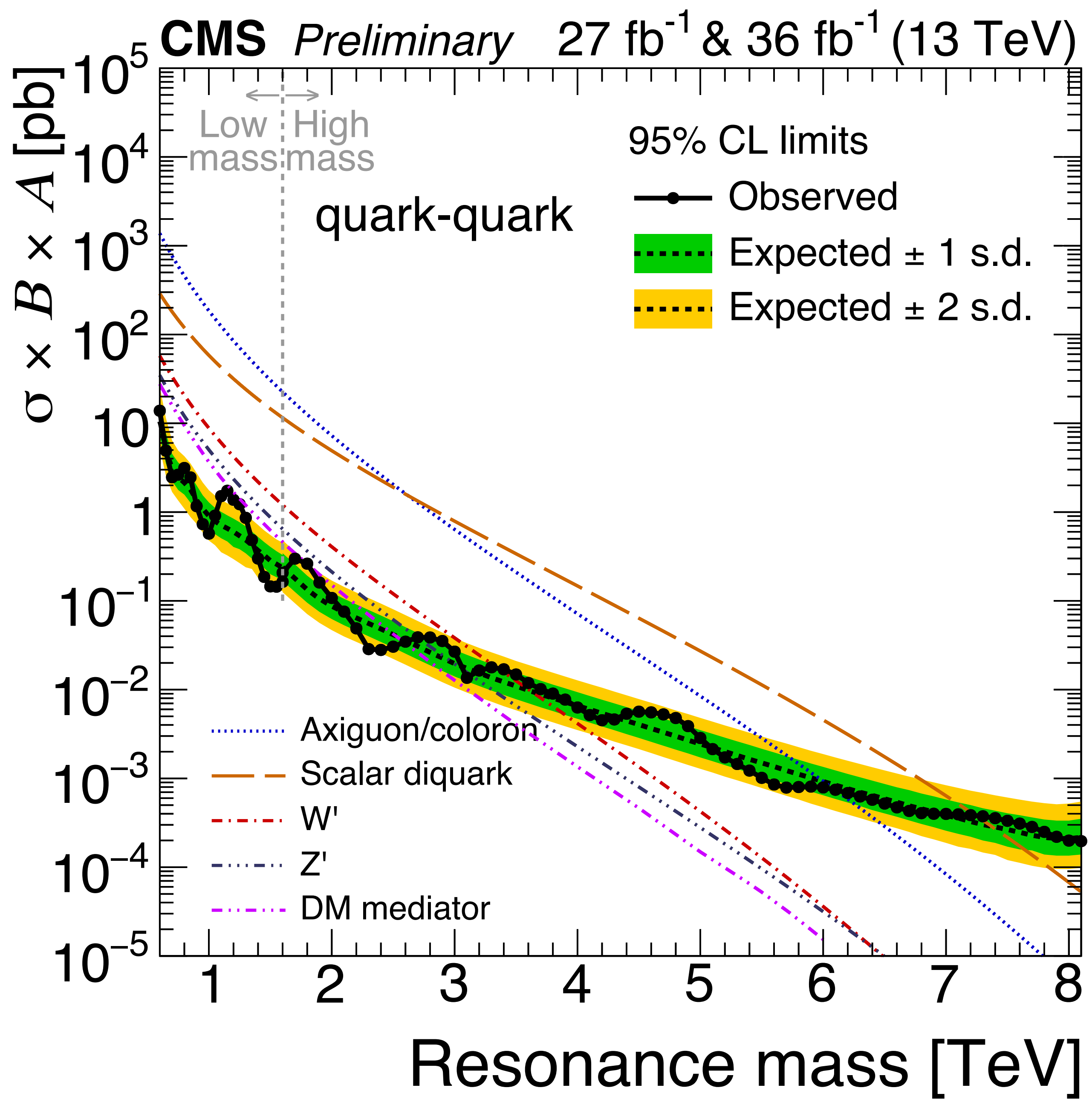
Their low mass analysis hasn't been updated since 3/fb.



Dijet resonances

ATLAS and CMS both presented dijet resonance searches at Moriond.

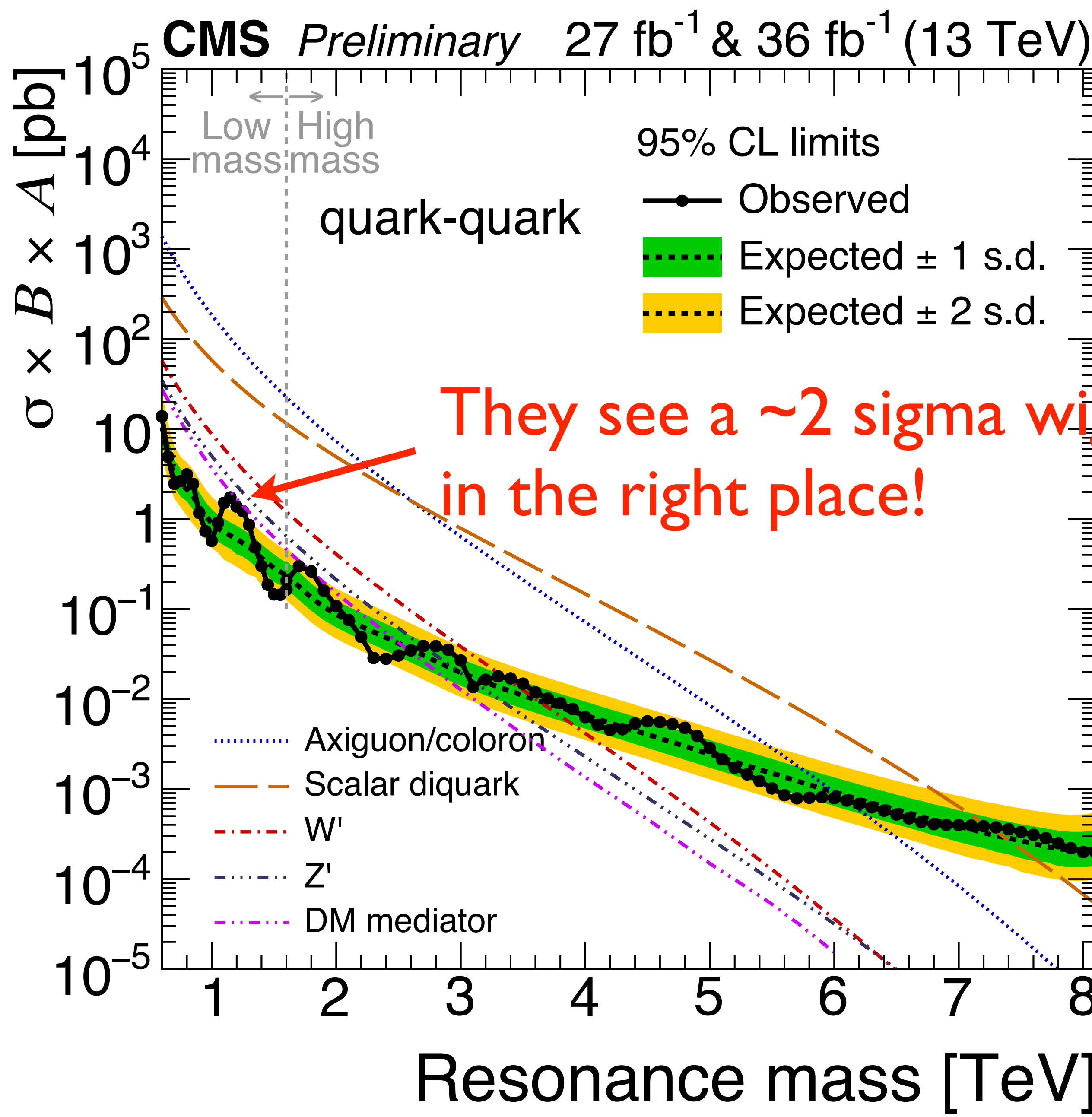
CMS covers the region below 1.6 TeV using a novel technique based on “scouting” data.



Dijet resonances

ATLAS and CMS both presented dijet resonance searches at Moriond.

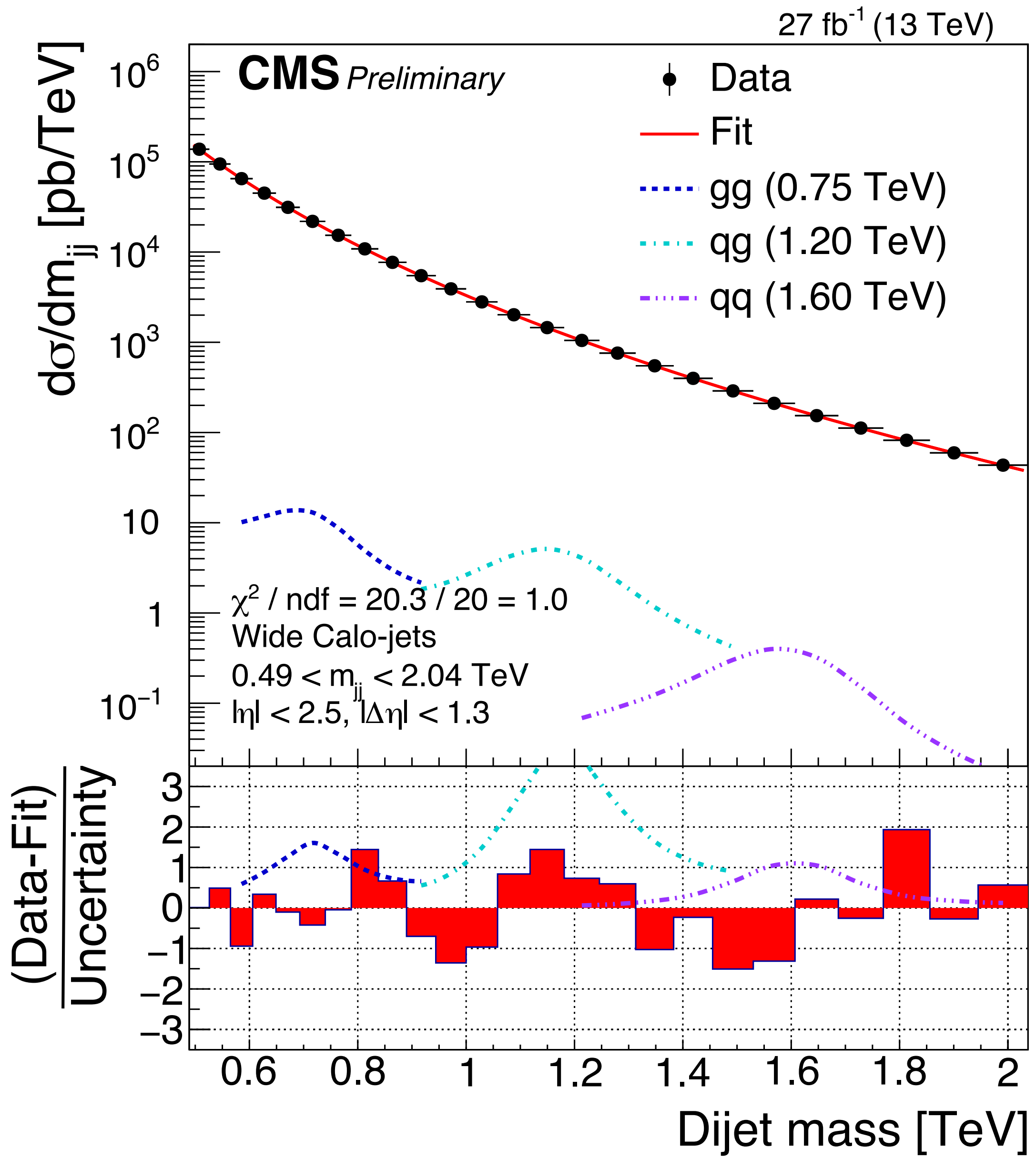
CMS covers the region below 1.6 TeV using a novel technique based on "scouting" data.



They see a ~2 sigma wiggle in the right place!

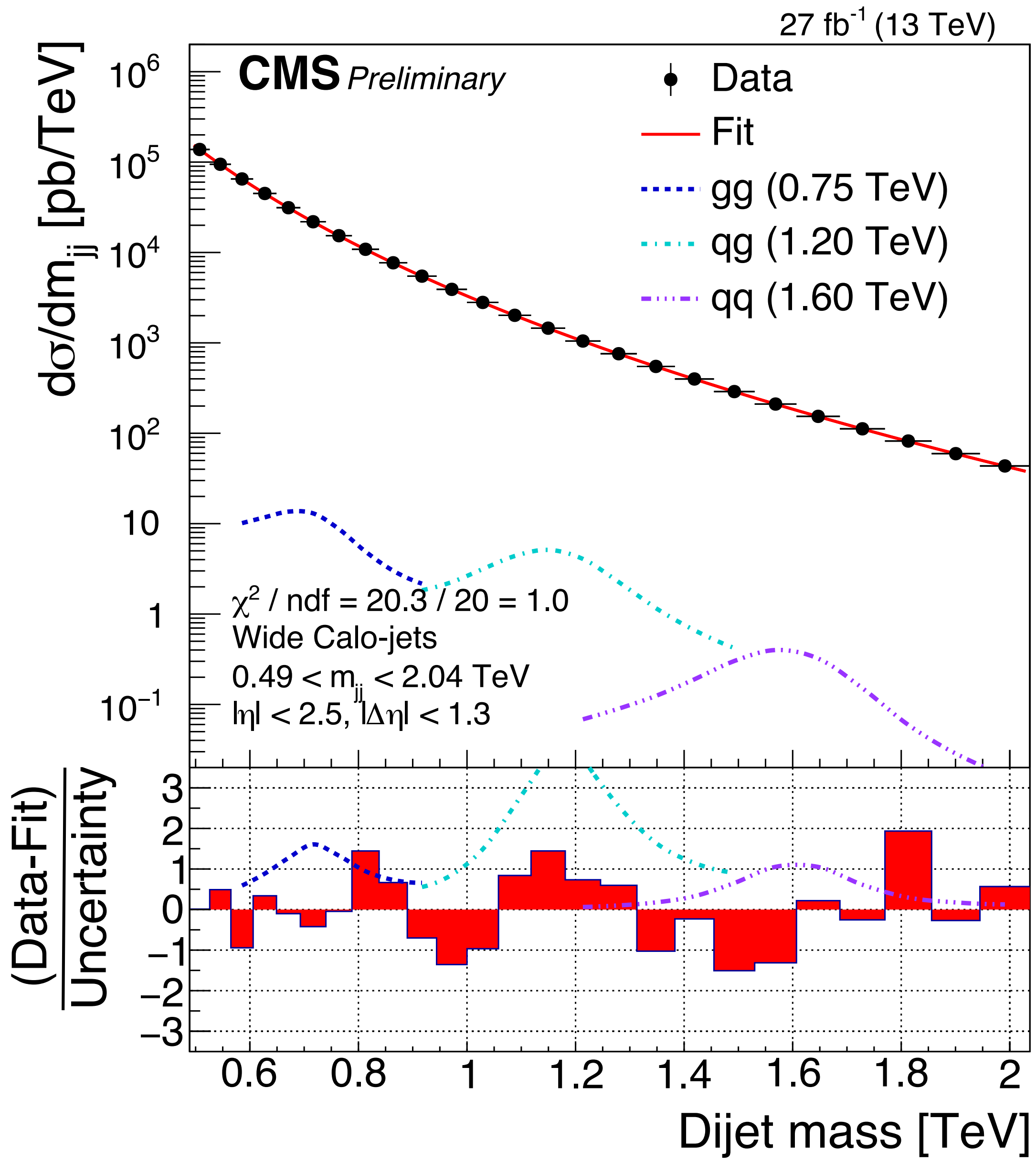
Dijet resonances

Unfortunately, CMS did not provide enough information for us to compute the statistical significance and add it to those from the jets+MET searches.



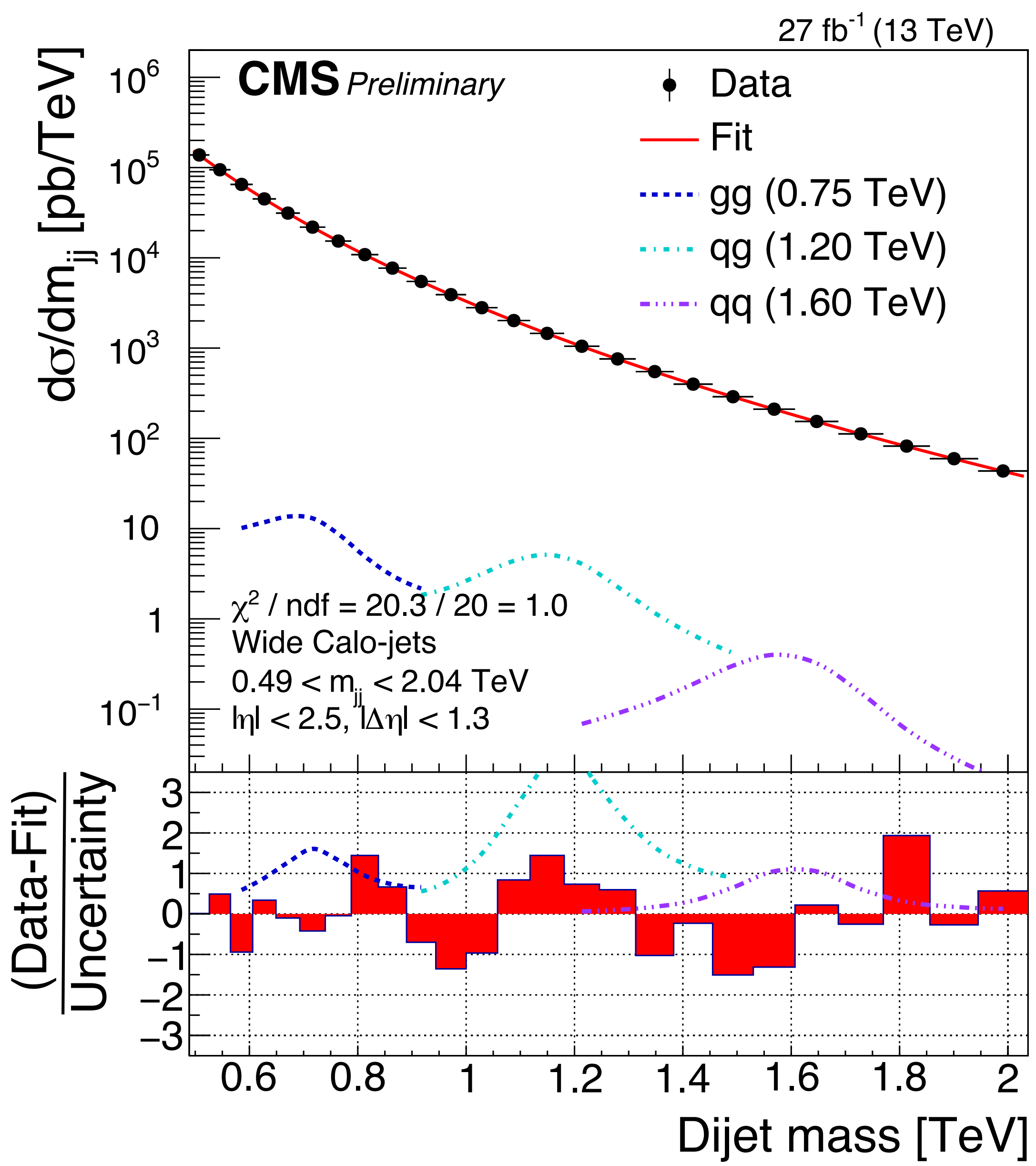
Dijet resonances

Unfortunately, CMS did not provide enough information for us to compute the statistical significance and add it to those from the jets+MET searches.



Dijet resonances

Unfortunately, CMS did not provide enough information for us to compute the statistical significance and add it to those from the jets+MET searches.



But one can dream....

Conclusions

We have performed an unprecedentedly deep dive into the SRs of the CMS jets+MET searches.

Using a simple method of "rectangular aggregations", we uncovered a number of potentially interesting excesses.

Exploring the most significant excess (3.5sigma, jets+MT2), we showed on model independent grounds that it is incompatible with jets+MHT.

The next-most-significant excess (3.2sigma, jets+MT2) proved to be very compatible with jets+MHT. It was well-fit with a "monosquark" model. It also had the potential to explain a ~2sigma wiggle in the CMS dijet resonance search.

Outlook

This is just the tip of the iceberg! There are several other excesses to explore in CMS jets+MET, and probably many more in the countless other searches for new physics.

In an era where the data doubling time will stretch to years, these excesses will stick around much longer. Simply waiting for more data becomes a less and less viable option.

The official propaganda plots based on a handful of simplified models just don't do justice to the richness and complexity of the data.

This could be LEE...or one of these excesses may turn out to be real. Do you want to be first or last?

Suggestions for experimentalists

Very useful (essential) information all analyses should provide to theorists:

- error correlation matrices
- digitized tables of the 100+ SRs in each search
- detailed cutflows and histograms for sample signal points
- enough information to calculate significance and fit models
  (CMS dijets, I'm looking at YOU)
- ....

It's important to explore many more simplified models, not necessarily MSSM inspired, in order to fully map out all the SRs of the CMS searches.

ATLAS should really consider adopting the CMS approach of fine binning and look at lower pTs. They are missing a lot!

Are there hints for new physics in the Moriond data?

Are there hints for new physics in the Moriond data?

Definitely maybe!

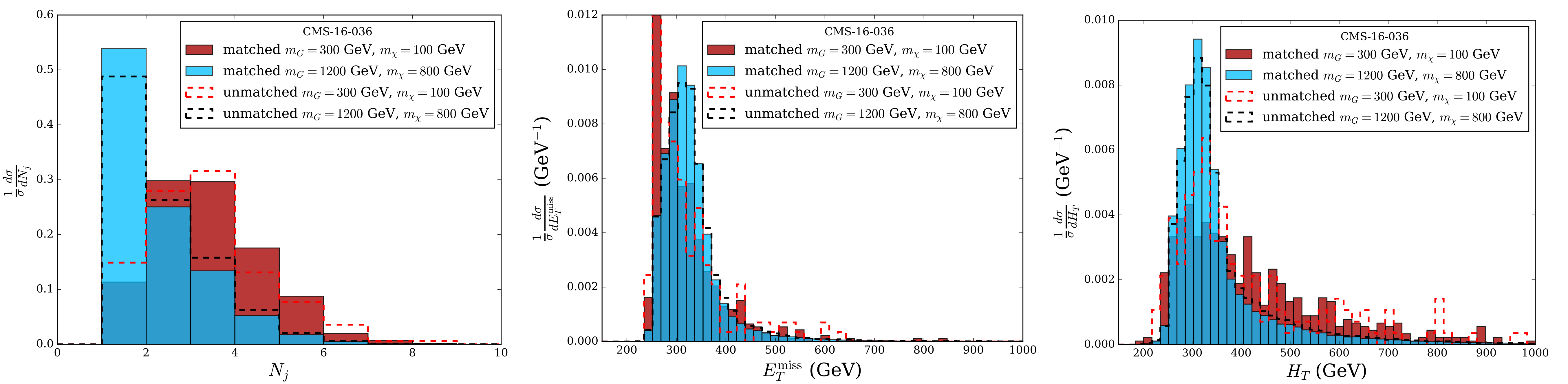
Thanks for your attention!

Backup material

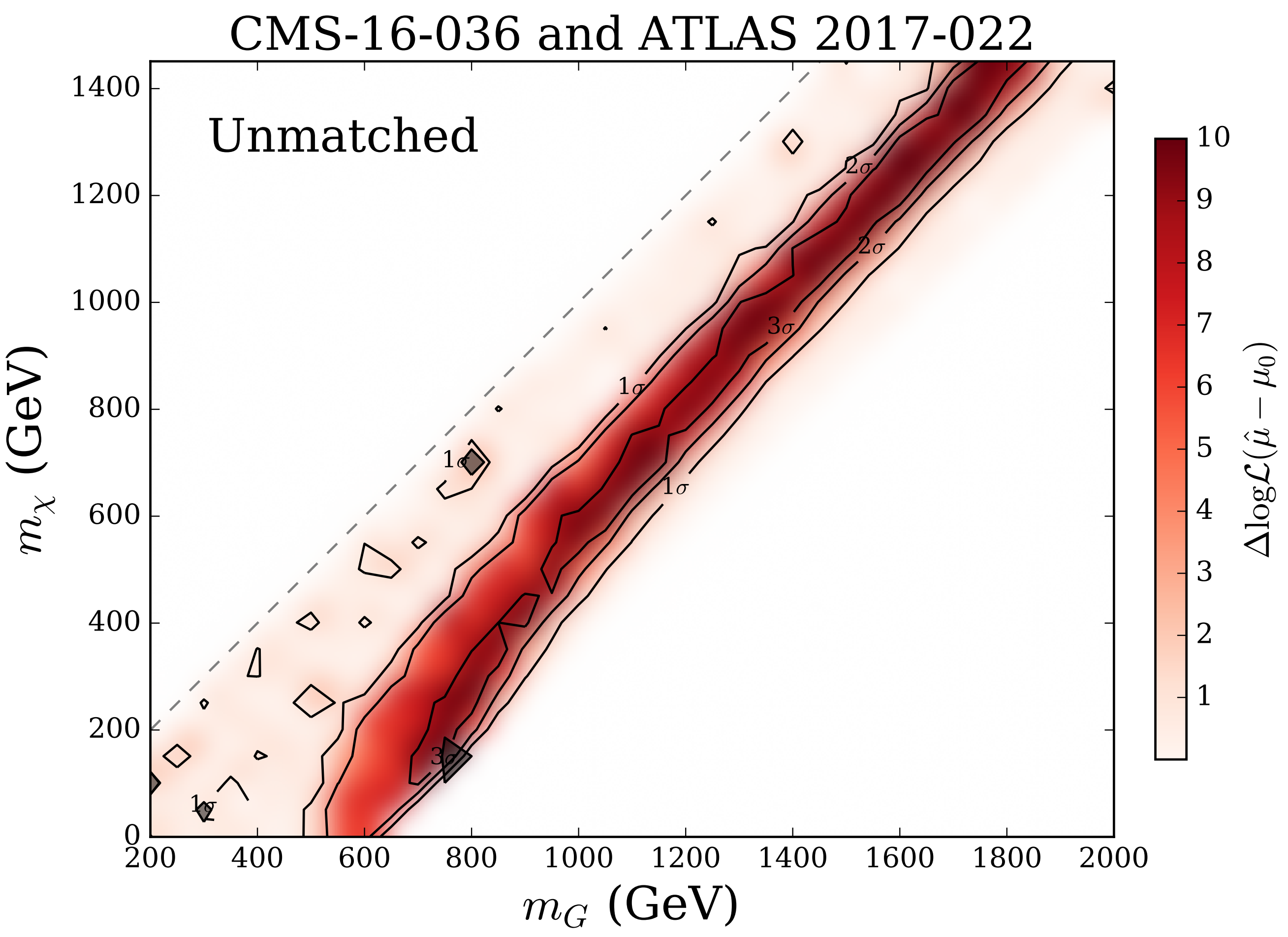
Comment on matching

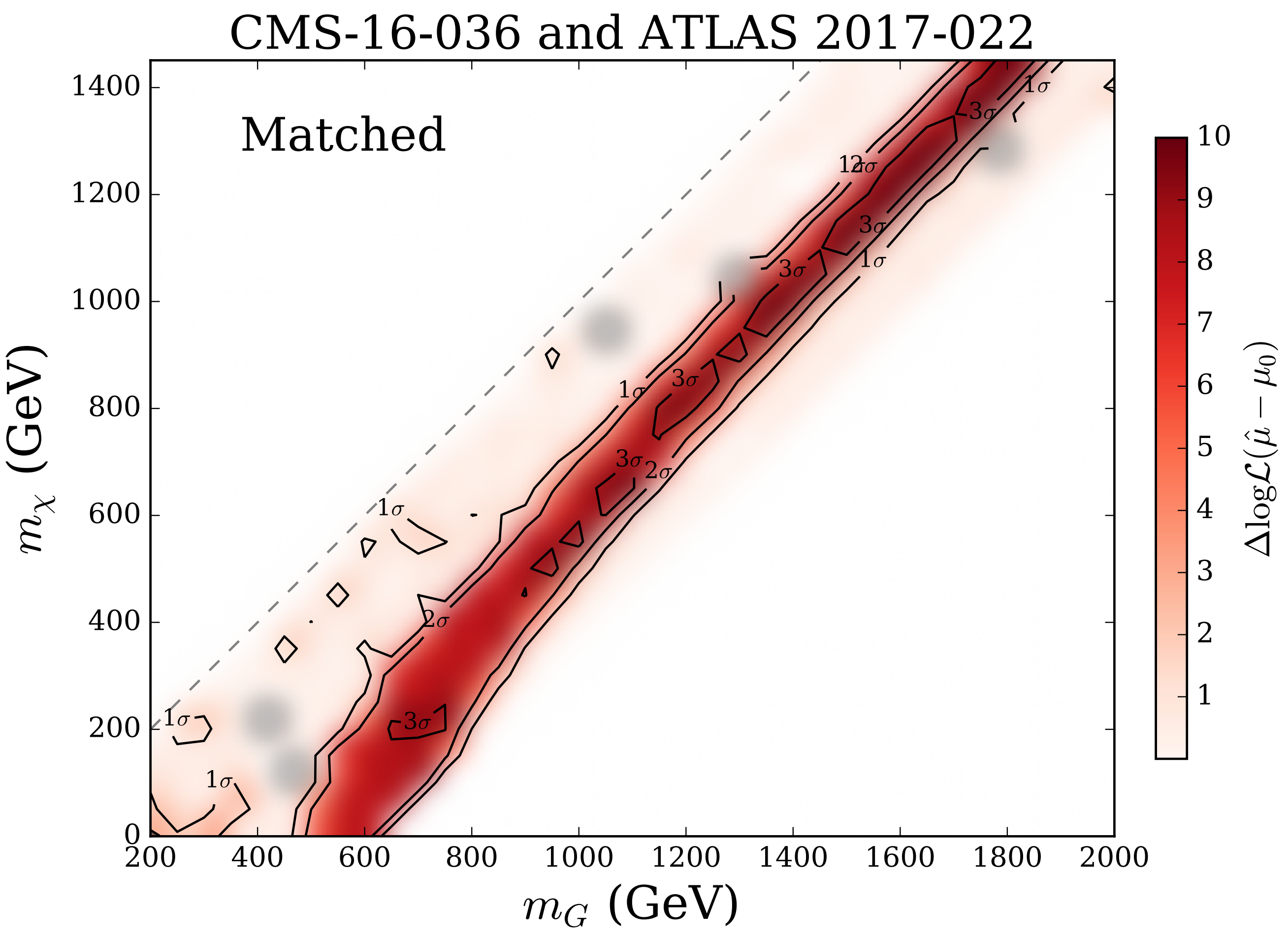
Unfortunately, Pythia8 has a bug in color flow through the UDD vertex that prevents us from producing matched samples of the monosquark model. We are in touch with the Pythia authors to address this.

But we checked that matching has a very minor effect in an analogous “sgluon” model (qqbar -> G -> g + chi).



Comment about matching





Effect of correlations

Consider the second most significant aggregation in 036. This is being driven by SRs 1,13

```
Out[645]= {{{1, Bin}, {250, 350, pT}, {1, 1, Nj}, {0, 0, Nb}, {0, ∞, MT2}, {63 253, 515, 509, 2677, Bkg}, {67 052, Data}},
  {{13, Bin}, {250, 450, HT}, {2, 3, Nj}, {0, 0, Nb}, {200, 300, MT2}, {60 324, 471, 467, 3228, Bkg}, {63 791, Data}}}
```

Without correlations, this is only 1.7 sigma.

With correlations it becomes 2.7 sigma!

Not every search provides a covariance matrix. These are essential, and all searches should provide them!!

CMS jets+MET searches

033 uses MHT -- negative vector sum of all jets in the event

036 uses MT2 (Lester & Summers 9906349): First cluster all jets into two hemispherical pseudojets, then compute, assuming massless invisible particles

$$M_{\mathrm{T2}} = \min_{\vec{p}_{\mathrm{T}}^{\mathrm{miss}\,X(1)} + \vec{p}_{\mathrm{T}}^{\mathrm{miss}\,X(2)} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}}} \left[\max\left(M_{\mathrm{T}}^{(1)}, M_{\mathrm{T}}^{(2)} \right) \right]$$

Key observation: when only two jets in the event, MT2 given by a simple analytical formula (Lester, 1103.5682):

$$M_{T2}^2 = 2(p_{T1} p_{T2} + \vec{p}_{T1} \cdot \vec{p}_{T2})$$

So MT2 in this analysis always obeys an inequality

$$M_{T2}^2 \leq MHT^2 = (\vec{p}_{T1} + \vec{p}_{T2})^2 = p_{T1}^2 + p_{T2}^2 + 2\vec{p}_{T1} \cdot \vec{p}_{T2}$$

Most significant RA

Let's consider the most significant RA in our list: 3.5 sigma from 036, corresponding to 1000<HT<1500; Nj=2,3; Nb=0,1; MT2>400.

It's driven largely by just two SRs (3 sigma):

```
Out[648]= {{{126, Bin}, {1000, 1500, HT}, {2, 3, Nj}, {0, 0, Nb}, {400, 600, MT2}, {110, 8, 7, 18, Bkg}, {159, Data}},
  {{127, Bin}, {1000, 1500, HT}, {2, 3, Nj}, {0, 0, Nb}, {600, 800, MT2}, {40, 3, 3, 7, Bkg}, {64, Data}}}
```

Using our MT2<MHT inequality, we can map these onto the following SRs of 033:

```
Out[649]= {{{6, Bin}, {350, 500, HTmiss}, {1000, ∞, HT}, {2, 2, Nj}, {0, 0, Nb}, {122.1, 9.5, 8.6, 8.8, 7.6, Background}, {139, Data}},
  {{8, Bin}, {500, 750, HTmiss}, {1000, ∞, HT}, {2, 2, Nj}, {0, 0, Nb}, {77.3, 6.8, 5.7, 6.1, 5.4, Background}, {96, Data}},
  {{9, Bin}, {750, ∞, HTmiss}, {750, 1500, HT}, {2, 2, Nj}, {0, 0, Nb}, {330, 13, 42, 12, 38, Background}, {272, Data}},
  {{36, Bin}, {350, 500, HTmiss}, {1000, ∞, HT}, {3, 4, Nj}, {0, 0, Nb}, {506, 18, 30, 17, 26, Background}, {490, Data}},
  {{38, Bin}, {500, 750, HTmiss}, {1000, ∞, HT}, {3, 4, Nj}, {0, 0, Nb}, {284, 12, 17, 12, 16, Background}, {303, Data}},
  {{39, Bin}, {750, ∞, HTmiss}, {750, 1500, HT}, {3, 4, Nj}, {0, 0, Nb}, {373, 14, 44, 13, 41, Background}, {334, Data}}}
```

There's nothing going on here. In fact, any signal that would explain all of 036 is ruled out by 033 at the 95% CL!

Second most significant RA

Let's move on to the next most significant RA: 3.26 sigma also from 036, corresponding to Nj=1-3; Nb=0-1; 250<HT<450; 200<MT2<300

This is driven largely by these two SRs (2.74 sigma):

```
Out[651]= {{{1, Bin}, {250, 350, pT}, {1, 1, Nj}, {0, 0, Nb}, {0, ∞, MT2}, {63 253, 515, 509, 2677, Bkg}, {67 052, Data}},
  {{13, Bin}, {250, 450, HT}, {2, 3, Nj}, {0, 0, Nb}, {200, 300, MT2}, {60 324, 471, 467, 3228, Bkg}, {63 791, Data}}}
```

These map onto these SRs of 033:

```
{{{1, Bin}, {300, 350, HTmiss}, {300, 500, HT}, {2, 2, Nj}, {0, 0, Nb}, {20 370, 120, 980, 120, 960, Background}, {21 626, Data}},
 {{4, Bin}, {350, 500, HTmiss}, {350, 500, HT}, {2, 2, Nj}, {0, 0, Nb}, {13 076, 93, 630, 93, 620, Background}, {14 019, Data}},
 {{31, Bin}, {300, 350, HTmiss}, {300, 500, HT}, {3, 4, Nj}, {0, 0, Nb}, {13 608, 110, 560, 110, 540, Background}, {14 520, Data}},
 {{34, Bin}, {350, 500, HTmiss}, {350, 500, HT}, {3, 4, Nj}, {0, 0, Nb}, {7496, 70, 330, 69, 320, Background}, {7973, Data}}}
```

These are also 2.6 sigma significant! Very compatible!

(Note, 033 lacks the Nj=1 bins.)

Distinctive jet pT distributions