An NGS-based carrier screen for Congenital Adrenal Hyperplasia with 95% detection rate

Dale Muzzey, Mark R. Theilmann, Kevin M. D'Auria, Henry H. Lai, Clement S. Chu, Imran S. Haque, Eric A. Evans, H. Peter Kang, & Jared R. Maguire

Counsyl

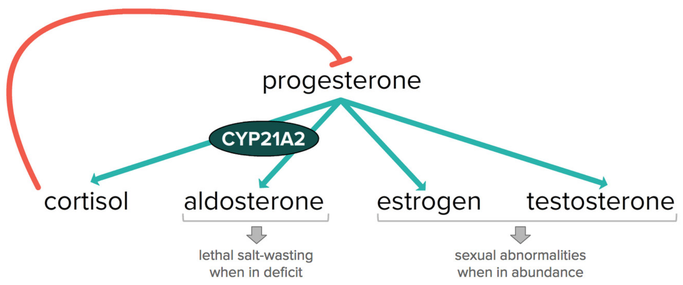
All Counsyl posters available online at **research.counsyl.com**

Carrier status of one of the ten most-common recessive diseases — generally absent from carrier screens due to highly homologous pseudogene — determined via custom NGS assay.

Abstract

There are two strong arguments for including congenital adrenal hyperplasia (CAH) on expanded carrier screen (ECS) panels: (1) with a carrier rate of 1 in 60, it is one of the ten most common recessive diseases, and (2) screening only ten variants in *CYP21A2* affords a 95% detection rate. However, CAH is absent from most ECS panels — or at best detected at a rate <20% — because 99% homology between the coding sequences of *CYP21A2* and its pseudogene *CYP21A1P* complicates variant identification. Indeed, eight of the ten common variants are pseudogene-derived; thus, expensive, low-throughput assays (e.g., long-range PCR + Sanger sequencing) are typically used to assess CAH carrier status. Here we report the development of an NGS-based carrier screen for CAH that detects all ten common deleterious variants in *CYP21A2*. The screen has a detection rate of 95% and has been applied to more than 100,000 patient samples.

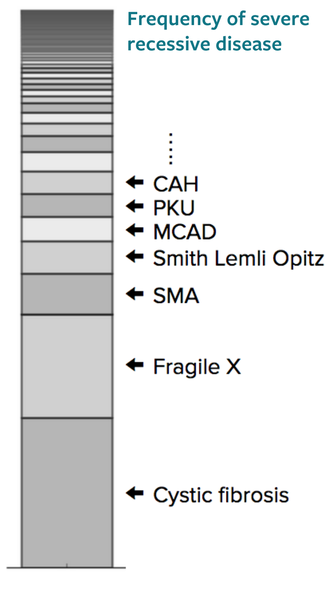
What is CAH?



- Severe recessive disease that affects steroidogenesis
- 95% of CAH caused by mutations in *CYP21A2*

Why test for CAH?

- CAH is 7th most-common recessive disease
- Carrier rate = 1 in 60; affected rate = 1 in 15000
- High false-positive rate on newborn screening
- Despite high prevalence, CAH is missing from nearly all commercial expanded carrier screens
- When included, carrier couples detected with rate of only 9%



Why is CAH screening hard?

- *CYP21A2* and pseudogene *CYP21A1P* have 99% identical coding regions
- 8 of the 15 distinguishing sites are deleterious
- Biological challenge: Rearrangements are frequent
- Technical challenge: Read mapping is difficult and CNV calling required.

Variant calling strategy

Exploit two forms of evidence:

CNV-style evidence

- Look for cases where CN(gene@site) < CN(gene)
  - These signal where pseudogene-derived sequence transferred into gene
  - Required development of technique to make confident single-base CN calls

Ex:

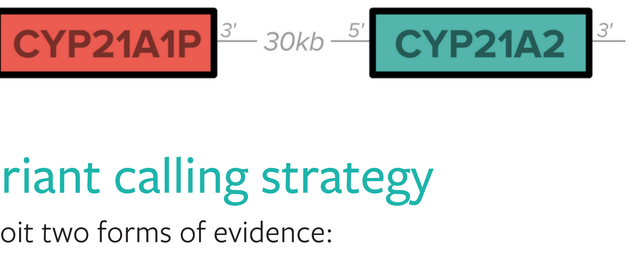
- CN(gene) = # of *CYP21A2* "whole" genes = 2
- CN(gene@site) = # of gene-derived bases @ I173 = 1

→ CN(gene@site) < CN(gene) → Carrier for I173N

SNP-style evidence

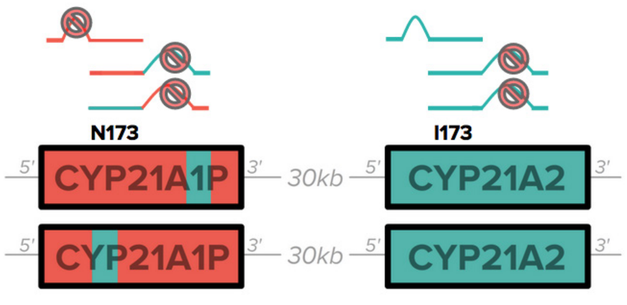
- From gene-derived reads, look for pseudogene-derived bases @ site of interest

Ex:



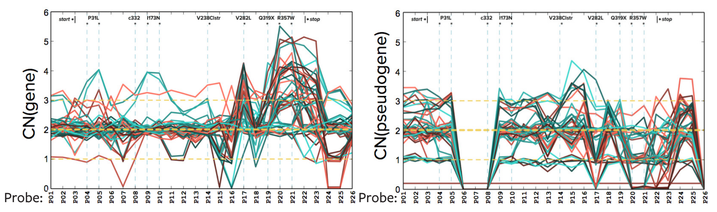
- Use CNV-calling to exclude probes with gene-derived sequence at one end if that sequence has transferred into pseudogene, because other end may trivially contribute pseudogene-derived reads

Ex:

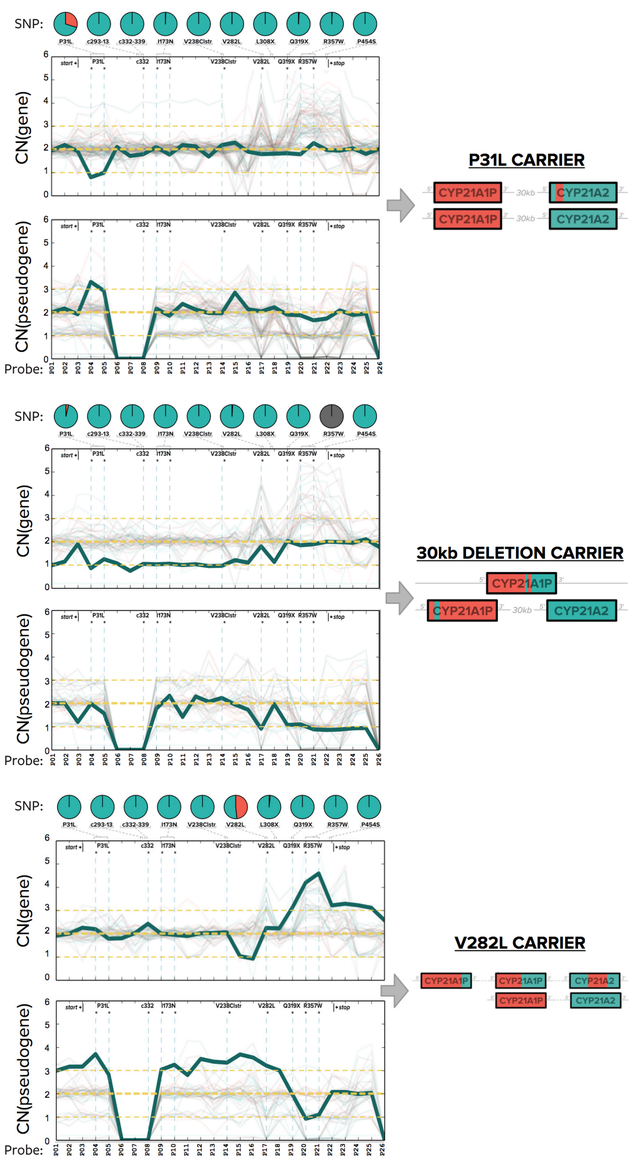


Extensive recombination

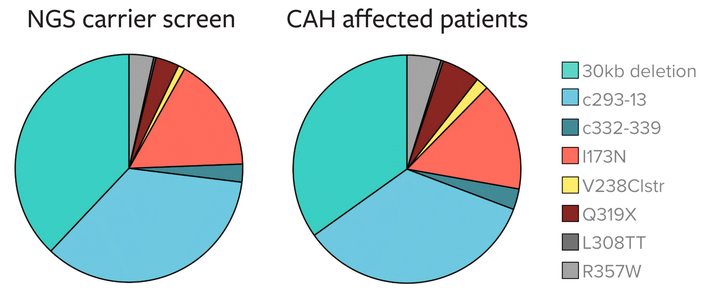
- Data from production batch of 96 samples reveals diversity of haplotype copy number and composition.



Sample data



Screen matches patient data



Conclusion

With the appropriate probe design and analysis suite, CAH can be assessed with 95% detection rate on NGS-based ECS panels. We have screened more than 100,000 patient samples for CAH and find variant frequencies consistent with data from affected patients.