

# Abracadabra

Finding genetic mutations in Go



Vitor De Mario  
[twitter.com/vdemario](https://twitter.com/vdemario)  
[github.com/vdemario](https://github.com/vdemario)

# Mendelics

- Mendelics is a genomic analysis lab from São Paulo, Brazil.
- We do genetic tests, trying to find mutations that explain diseases on our patients so that they can get treatment.

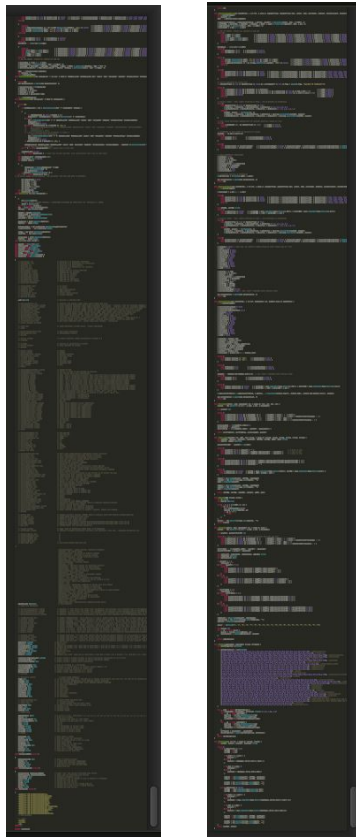
# Abracadabra

- Each patient has thousands of mutations.
- We needed a fast and accurate way to point out which ones caused diseases.
- Abracadabra is built in Go and uses Random Forests to solve that problem.

## The beginning

- CEO, a neurologist, started programming by himself.
- He chose Go almost an accident.
- A prototype was built despite the lack of experience.

# The monolith



## Learning Go

- No one on the team had experience with Go before.
- In just a few weeks we were able to get a good grasp of the language and started becoming productive.
- Great documentation.

## Growing pains

- Bad habits from other languages.
- We loaded the whole genome in memory, it wasn't a good implementation at first
  - 3+ billion bases (ACTG)
  - 3 gigabases == gigabytes
- *go tool pprof* to the rescue



## Moving forward

- The new language was no longer an issue
- In less than a month we could focus on our real issues:
  - bioinformatics, genetics and machine learning



# Open source: github.com/mendelics/vcf

[GoDoc](#) [Home](#) [Index](#) [About](#)

[vcf: github.com/mendelics/vcf](#) [Index](#) | [Examples](#) | [Files](#)

## package vcf

```
import "github.com/mendelics/vcf"
```

Package vcf provides an API for parsing genomic data compliant with the Variant Call Format 4.2 Specification

This API is built with channels, assuming asynchronous computation. Variants parsed successfully are sent immediately to the consumer of the API through a channel, as well as variants that fail to be processed.

Example

## Index

```
func SampleIDs(reader io.Reader) ([]string, error)
func ToChannel(reader io.Reader, output chan<- *Variant, invalids chan<- InvalidLine) error
type InvalidLine
type SVType
    func (i SVType) String() string
type Variant
    func (v *Variant) String() string
```

# Machine Learning in Go

- RandomForest<sup>TM</sup>
- [github.com/ryanbressler/CloudForest](https://github.com/ryanbressler/CloudForest)

## CloudForest

---

build passing

godoc reference

[Google Group](#)

Fast, flexible, multi-threaded ensembles of decision trees for machine learning in pure Go (golang).

# Machine Learning in Go

- Trial and error
- CloudForest is powerful



# Machine Learning in Go

- Over 90% accuracy
- Predicting a mutation is benign is much more precise. Thousands of mutations are filtered, avoiding waste of time from our geneticists.

# Today

- Every patient goes through Abracadabra.
- More than two years in production, users refuse to go back.
- Moving towards TensorFlow.

# Today

HERANÇA COMPATÍVEL

HERANÇA INCOMPATÍVEL

VARIANTES NÃO RELEVANTES

BUSCA DE GENES

1 - BRCA2 (Breast Cancer 2, Early Onset)

IGV

READS

PUBMED

 REMOVER DO LAUDO

- [612555] Familial Breast-Ovarian Cancer (BROVCA2) | Herança: monoallelic
- [613029] Glioma Susceptibility (GLM3) | Herança: monoallelic
- [613347] Pancreatic Cancer | Herança: monoallelic
- [114480] Breast Cancer | Herança: monoallelic
- [155255] Medulloblastoma (MDB) | Herança: monoallelic
- [605724] Fanconi Anemia (FANCD1) | Herança: biallelic

Variante	UCSC	IGV	Cópias	Efeito	MAF	Clinvar	PubMed	Predição	Consenso	Qual.	Validação
Chr13:32,905,109-32,905,110 Ref: AT (2bp) Alt: A	UCSC	IGV	1	Frameshift p.Phe246Leu fs*5 AA 246/3418, Exon 9/27 ENST00000380152	0.000000			Deletério	pathogenic ⓘ		

## What Go made possible

- Simple enough for a physician to start.
- Powerful enough to do machine learning in production.
- Most of our problems solved in the stdlib.
- All of that built in Brazil.

GopherCon Brasil - [gopherconbr.org/en](http://gopherconbr.org/en)





# Thanks



[twitter.com/vdemario](https://twitter.com/vdemario)

[github.com/vdemario](https://github.com/vdemario)

